

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2004, 10:50:55 ; Search time 16521.2 Seconds  
(without alignments)  
11538.043 Million cell updates/sec

Title: US-10-022-710-1  
Perfect score: 4398  
Sequence: 1 atggtgcgtgcattcagaa.....atggagacttagacatgtaa 4398

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.ov.\*
- 5: gb.pat.\*
- 6: gb.ph.\*
- 7: gb.pl.\*
- 8: gb.pr.\*
- 9: gb.ro.\*
- 10: gb.sts.\*
- 11: gb.vi.\*
- 12: gb.vi.\*
- 13: gb.vi.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.mu.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg.hum.\*
- 31: em.htg.inv.\*
- 32: em.htg.other.\*
- 33: em.htg.mus.\*
- 34: em.htg.pin.\*
- 35: em.htg.ram.\*
- 36: em.htg.ram.\*
- 37: em.htg.vrt.\*
- 38: em.sv.\*
- 39: em.hgo.hum.\*
- 40: em.hgo.mus.\*
- 41: em.hgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	4380.4	99.6	5724	9	AB051466 Homo sapi
2	3321.2	75.5	6112	10	BC063250 Mus muscu
3	2200.2	50.0	2421	9	BC033125 Homo sapi
4	2200.2	50.0	2422	9	BC019344 Homo sapi
5	1232.8	28.0	5669	9	AB023177 Homo sapi
6	1232.8	28.0	6373	6	AX079870 Sequence
7	1229.6	28.0	6378	6	AR338841 Sequence
8	1129.6	25.7	5032	6	AX818221 Sequence
9	1109.8	25.2	5287	6	AX780247 Sequence
10	580.6	13.2	3053	6	AX747433 Sequence
11	580.6	13.2	3053	9	AK092252 Homo sapi
12	527	12.0	119205	9	AC011231 Homo sapi
13	527	12.0	167462	2	AC027600 Homo sapi
14	362.4	8.2	224789	2	AC111611 Rattus no
15	360.8	8.2	259236	2	AC123741 Mus muscu
16	270.4	6.1	169928	9	AC019067 Homo sapi
17	239.2	5.4	194472	9	AC012000 Homo sapi
18	198.4	4.5	225453	2	AC097957 Rattus no
19	196.4	4.5	89212	9	AC016679 Homo sapi
20	189.2	4.3	215703	2	AC121276 Mus muscu
21	175.8	4.0	68894	2	AC100201 Mus muscu
22	166.4	3.8	258545	2	AC110536 Mus muscu
23	165	3.8	259236	2	AC123741 Mus muscu
24	163.4	3.7	277376	2	AC128972 Rattus no
25	154	3.5	62063	2	AC136706 Mus muscu
26	154	3.5	183783	9	AC017082 Homo sapi
27	137.6	3.1	256545	2	AC110536 Mus muscu
28	133.8	3.0	399	11	G61408 SHGC-85947
29	132.2	3.0	198944	2	AC115912 Mus muscu
30	132	3.0	221262	2	AC122394 Mus muscu
31	128.8	2.9	872	11	BV023652 S212P6059
32	124.2	2.8	95921	9	AC004141 Homo sapi
33	124.2	2.8	168008	9	AC142334 Pan trogl
34	124	2.8	215510	2	AC128337 Rattus no
35	124	2.8	263244	2	AC097540 Rattus no
36	119.4	2.7	198944	2	AC115912 Mus muscu
37	117.6	2.7	110000	2	Continuation (4 of
38	117.6	2.7	110000	2	Continuation (5 of
39	116.8	2.7	256124	2	AC130447 Rattus no
40	108.8	2.5	191439	10	AL627302 Mouse DNA
41	108.8	2.5	199140	2	BX005251 Mus muscu
42	103.6	2.4	1704	9	BC016980 Homo sapi
43	102	2.3	300201	2	AC107086 Rattus no
44	100.2	2.3	143751	9	AC004160 Homo sapi
45	98.2	2.2	85693	2	AC124117 Mus muscu

ALIGNMENTS

RESULT 1  
AB051466  
LOCUS  
DEFINITION Homo sapiens mRNA for KIAA1679 protein, partial cds.  
ACCESSION AB051466  
VERSION AB051466.1 GI:12697902  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (sites)  
Nagase, I., Kikuno, R., Hattori, A., Kondo, Y., Okumura, K. and Ohara, O.  
Prediction of the coding sequences of unidentified human genes.  
XIX. The complete sequences of 100 new cDNA clones from brain which

code for large proteins in vitro  
DNA Res. 7 (6), 347-355 (2000)  
21082932

11214970  
2 (bases 1 to 5724)  
Chara, O., Nagase, T. and Kikuno, R.  
Direct Submission  
Submitted (22-NOV-2000) Osamu Ohara, Kazusa DNA Research Institute,  
Department of Human Gene Research; 1532-3, Yana, Kisarazu, Chiba  
292-0812, Japan (E-mail: cdnainfo@kazusa.or.jp,  
URL: http://www.kazusa.or.jp/huge, Tel: 81-438-52-3913,  
Fax: 81-438-52-3914)

## FEATURES

source

1. 5724  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="fh22954"  
/notes="vector:pBluescriptII SK plus"  
1. 5724  
/genes="KIAA1679"  
1. 5724  
/genes="KIAA1679"  
1. 5724  
/notes="Start codon is not identified."  
/codon\_start=2  
/product="KIAA1679 protein"  
/protein\_id="BAB21770.1"  
/db\_xref="GI:12697903"  
/translation="GNTSLNSGNSRPPKRSFRVCDWSDLFQWVSDWVHCVL  
VPVARGVKPRTEACVTAQGLHRVVRICIKLNRTVANE:CEHFALQPTQACLI  
PCPRCDVSVFLPWSNCKGKLOHRTAVIAPLFGLOCPNLTESRACDAPLS  
PLGEEVTFSLKVPNSKCLPHLKEINPSRGTVDPSNERTVFKHOSKAKHHS  
KSWAIEGYTQVSCRSQGNAMLSLCLQDSFPLTVOSKMPKDCETSONKSPC  
SKTSCRSGLPGRSRNRVKNVAIGGCKECPLEKEACIVVEGELLOCPKRYSWETS  
EWKESQSLLEQDPHWHVTGCVGGGIQREVIQAQSVPAALAAAEKSRVEXKA  
LCVGPAPLPQCLNIPGSDICVSSWSAWGLCHENHCEPQKGGKFRTRVHLMST  
GPAGHCPLHVESVPCEDPCVYVLAISGICFPDHRKCGHLRLKAVQCDNDEGVSG  
SLCPVPPPKRSCEIICPRMDCLSEWTEWSQSCSNKNSDKGRTSRITIALAGE  
GKPCPSQALQHLRCHNDHSQMLWETSPGFCSEDTLVTALNATIWNGEATCGV  
GIOTRVKSVSHVQVWTKRCPDSTPEVTRVPCFLPKCKDCIVTAFSEWTPCPRMCO  
AGNAVQSRRIIIOAANGGCEPDTLVEEDECEDVSLCPVYRKKPKQWSPCLIVP  
ESVWQGITSGSEAGLQTRAVESCISDDNRSAEMMECLQTNMGMLLVQECTVPRE  
DCTFTAWSKTFCPCNEATKRRRLQTKSRKKEQKQSDSLYLVETELCPCEFTIS  
QPVNWSKDLIPERREGLRVLQADSKCEGLRFAVACDKNRPVDFPSCSS  
GYIEKCVIPCPFKLCSWSGSSGSCSIGVIRSKWLKPKYNGRPPFKLDLK  
NOACVHEAPVCSYCNYSWVVEHWSCKINNELSLRCGGTOSKIRCVNTADGEG  
GAVDSNLQNDIEIPETQSCSLMPCNEVSEWLSKCPQSCDPHTMORRTHLLRP  
SLNSRTCAEDSQVPCLLNENCFQNLNFWSTQLSNAPQCGQVTRLLSCVCSLD  
GKPYMDQCSQHNLEKPFQMSIPCLVECVNQLSGWTAWTECSQTCGHHGRSRTF  
IIMPTQGEGRPCPTLTQKTCPTVPCYVNLGWSACKLEGGDCGEGVQIRLSLQV  
HSGISHAAGRVEDALGEMFPQDSILKQLCSVPQGDCHLTWSEWSTCELTIDGR  
SFETVGRQSRRTFIIOSFENQDSPOOVLTPCTCGKCYHYTWKASLWNNERTFW  
CORSDGVNVTGCSQSPARPAIROCIACRKPSPYCTQGGVCGCEKGYTEIMSNGL  
DYCMKVGSDEKADKXNLSKSRPNYSKXHDIFKXWSLQPLDQDGRVXWYGVSGG  
AFLIMIFLIFTSYLVCKKPKHQSPFPQCKPLILAYDGLDM"

gene

CDS

## ORIGIN

Query Match 99.6%; Score 4380.4; DB 9; Length 5724;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 4397; Conservative 0; Mismatches 1; Indels 6; Gaps 1;  
QY 1 ATGGTGGCTGCAATTCAGAGTGAACCGAATCTGGTGTGCAATGAAATATGGAACAC 60  
DB 209 ATGGTGGCTGCAATTCAGAGTGAACCGAATCTGGTGTGCAATGAAATATGGAACAC 268  
QY 61 TTTGGCTTTCAGCTTACAGACAGAGCTTGCCTCATCTTCTTCCCGGATTTGTGA 120  
DB 269 TTTGGCTTTCAGCTTACAGACAGAGCTTGCCTCATCTTCTTCCCGGATTTGTGA 328  
QY 121 GTATCTGAGTTCTTACCATGTCCTCAACTGTAGCAAGGAGTGGGAAGAAATTCAGCAT 180  
DB 329 GTATCTGAGTTCTTACCATGTCCTCAACTGTAGCAAGGAGTGGGAAGAAATTCAGCAT 388

QY 181 AGAATCGCGCGGTATAGCTCCCTCTCTTTGGTGGTTTGCATGTCCAAATCTGACT 240  
DB 389 AGAATCGCGCGGTATAGCTCCCTCTCTTTGGTGGTTTGCATGTCCAAATCTGACT 448  
QY 241 GAGTCAAGAGCCTGTGATGCTCCCATTTTCTGTCTCTTGGGGAAGAGGAATATACATTT 300  
DB 449 GAGTCAAGAGCCTGTGATGCTCCCATTTTCTGTCTCTTGGGGAAGAGGAATATACATTT 508  
QY 301 AGCCTTAAGTTGGACCATCGAGTAATGACAGACTGCTCATCTTAAAGAAATTAATCCA 360  
DB 509 AGCCTTAAGTTGGACCATCGAGTAATGACAGACTGCTCATCTTAAAGAAATTAATCCA 568  
QY 361 AGCGGAAGAACTGTTCTGGATTTTAACTCTGATTCAAATGAGCGAGTCACTTTAAACAT 420  
DB 569 AGCGGAAGAACTGTTCTGGATTTTAACTCTGATTCAAATGAGCGAGTCACTTTAAACAT 628  
QY 421 CAAAGTTTACAAAGCACATCATCTTGAAGTCTTGGCAATAGATAGTGTATCAAAACC 480  
DB 629 CAAAGTTTACAAAGCACATCATCTTGAAGTCTTGGCAATAGATAGTGTATCAAAACC 688  
QY 481 CGGCAAGTTTGTGTACAAGAGTGTGGAACAAATGCTATGTTAAAGCCTTTGCTTCAA 540  
DB 689 CGGCAAGTTTGTGTACAAGAGTGTGGAACAAATGCTATGTTAAAGCCTTTGCTTCAA 748  
QY 541 GATTCTCTCCCATTTGACTGTTTCACTCTGATCATGCCCCAAAGACTGTGAAACCTCCAG 600  
DB 749 GATTCTCTCCCATTTGACTGTTTCACTCTGATCATGCCCCAAAGACTGTGAAACCTCCAG 808  
QY 601 TGGTCTCTCTGGAGCCCTGCTCCAAAGACTGCTGCTGAGGAGTCTCTGCGCAGGATTT 660  
DB 809 TGGTCTCTCTGGAGCCCTGCTCCAAAGACTGCTGCTGAGGAGTCTCTGCGCAGGATTT 868  
QY 661 AGGAGCAGGAGCCGGAACGTTGAACACATGCTGATTTGGAGTGTGAAGAGTGTCTCGAA 720  
DB 869 AGGAGCAGGAGCCGGAACGTTGAACACATGCTGATTTGGAGTGTGAAGAGTGTCTCGAA 928  
QY 721 CTTCTTGAGAAAGAGGCTGCAATTTGTTGAAGAGAACTTCTGACAGCAATGTCCAGGAT 780  
DB 929 CTTCTTGAGAAAGAGGCTGCAATTTGTTGAAGAGAACTTCTGACAGCAATGTCCAGGAT 988  
QY 781 TCCTGGAGAACTTCTGAATGGAAGAAATGCAAGTCTCTCTCTCTCGAGCAGCAGGAT 840  
DB 989 TCCTGGAGAACTTCTGAATGGAAGAAATGCAAGTCTCTCTCTCTCGAGCAGCAGGAT 1048  
QY 841 CCCCCTGGCATGTGACGGGACCCGCTGTGTGGCGGTGGATTCAGACCCGGAGGTGTAC 900  
DB 1049 CCCCCTGGCATGTGACGGGACCCGCTGTGTGGCGGTGGATTCAGACCCGGAGGTGTAC 1108  
QY 901 TGTGCCAGAGGGTACAGCAGCTGCCGCACTGAGGCGCAAGGAAGTCTCTAGACCTGTG 960  
DB 1109 TGTGCCAGAGGGTACAGCAGCTGCCGCACTGAGGCGCAAGGAAGTCTCTAGACCTGTG 1168  
QY 961 GAAAAGGCAATATGTGTGGGACCCCGCTGTGCTCTGAGCTCTGCAATATCCCTTGC 1020  
DB 1169 GAAAAGGCAATATGTGTGGGACCCCGCTGTGCTCTGAGCTCTGCAATATCCCTTGC 1228  
QY 1021 TCTAGGAGTGTAGTATCTTCTGCTGCTGAGCTGCGGCTGTGCTCATGAAACTGT 1080  
DB 1229 TCTAGGAGTGTAGTATCTTCTGCTGCTGAGCTGCGGCTGTGCTCATGAAACTGT 1288  
QY 1081 CATGAACCTCAGGGGAAAAAGGATTTAGAACGAGGCGCAGCCCATGTCTCTCATGGAATCT 1140  
DB 1289 CATGAACCTCAGGGGAAAAAGGATTTAGAACGAGGCGCAGCCCATGTCTCTCATGGAATCT 1348  
QY 1141 ACAGGCTCTGAGGCGATGCTGCTCATTTGTTGGAGTCTGTTCTTGTGAGGATCCAATG 1200  
DB 1349 ACAGGCTCTGAGGCGATGCTGCTCATTTGTTGGAGTCTGTTCTTGTGAGGATCCAATG 1408  
QY 1201 TGTACCGATGCTGCGCATCAGAGGAGTCTGTTCTCTGATCATGGAAGAAATGTGCGCTG 1260  
DB 1409 TGTACCGATGCTGCGCATCAGAGGAGTCTGTTCTCTGATCATGGAAGAAATGTGCGCTG 1468  
QY 1261 GGACATCGTATTTCTGAAGGCGCTCTGCCAGATGACCGCGGAGAGATGTATCAGGAGT 1320



Db	1469	GGACATCGTATTCGAAAGCCGTCGCGAGAAATGACCGCGAGAAAGATGATCAGGAGT	1528
Qy	1321	CTTTGCCAGTTCCCCCTCTCTTGAGAGGAAGTCCTTGTAATTCCTCCGGAATGGAC	1380
Db	1529	CTTTGCCAGTTCCCCCTCTCTTGAGAGGAAGTCCTTGTAATTCCTCCGGAATGGAC	1588
Qy	1381	CTGTGCTGAGCGAGTGCACGAGTGGTCATCTCTGTTCCCACTCCTGTTCAAATATAAAAC	1440
Db	1589	TGTGTGCTGAGCGAGTGCACGAGTGGTCATCTCTGTTCCCACTCCTGTTCAAATATAAAAC	1648
Qy	1441	TCAGATGGGAAACAGACGAGTCAAGAACTATTCCTGGACATCGCTGGGGAAGTGGAAAG	1500
Db	1649	TCAGATGGGAAACAGACGAGTCAAGAACTATTCCTGGACATCGCTGGGGAAGTGGAAAG	1708
Qy	1501	CCATGTCCTCCCTAGTCAGGCTCTCCAAGAGACCTTTGTGTATGACCATTCCTCTGTATG	1560
Db	1709	CCATGTCCTCCCTAGTCAGGCTCTCCAAGAGACCTTTGTGTATGACCATTCCTCTGTATG	1768
Qy	1561	CAGCTTCACTGGGAGACATCGCCTTTGGGCCCCTTTCTGAGGACACATTCGTAACTGCCC	1620
Db	1769	CAGCTTCACTGGGAGACATCGCCTTTGGGCCCCTTTCTGAGGACACATTCGTAACTGCCC	1828
Qy	1621	CTTAATGCAACCATTCGCTGGAAATGGAAGACGACGTTGTGTAGGCAATTCAGACTCGG	1680
Db	1829	CTTAATGCAACCATTCGCTGGAAATGGAAGACGACGTTGTGTAGGCAATTCAGACTCGG	1888
Qy	1681	AGAGTCTTCTGTCTAAGAGTCACTGTGGGCAAGTAATGACCAAAAGATGTCAGATTCT	1740
Db	1889	AGAGTCTTCTGTCTAAGAGTCACTGTGGGCAAGTAATGACCAAAAGATGTCAGATTCT	1948
Qy	1741	ACTCGACCTGAAACTGTGGCCCCCTTTTCTCCCATGCAAAAAAGACTGTATTTGTGACT	1800
Db	1949	ACTCGACCTGAAACTGTGGCCCCCTTTTCTCCCATGCAAAAAAGACTGTATTTGTGACT	2008
Qy	1801	GCCTTCAGTGAAGTGGACACCTTCGCCCAAGGATGTGCCAAGCAGGAAATGCCACAGTAAAA	1860
Db	2009	GCCTTCAGTGAAGTGGACACCTTCGCCCAAGGATGTGCCAAGCAGGAAATGCCACAGTAAAA	2068
Qy	1861	CAGTCTCGATACAGAAATCATCTCAAGAAGCGCCATGGAAGGCCAGGAATGCCCGAT	1920
Db	2069	CAGTCTCGATACAGAAATCATCTCAAGAAGCGCCATGGAAGGCCAGGAATGCCCGAT	2128
Qy	1921	ACCTTATATCAGGAGAGAGTGTGAAGATGTTTCTTGTGTCTCTGTAATCGGTGGAG	1980
Db	2129	ACCTTATATCAGGAGAGAGTGTGAAGATGTTTCTTGTGTCTCTGTAATCGGTGGAG	2188
Qy	1981	CCACAGAAATGGAGCCCTTGCACTTTAGTGCACAGATCTGTCTGGCAGGGAATACGGGC	2040
Db	2189	CCACAGAAATGGAGCCCTTGCACTTTAGTGCACAGATCTGTCTGGCAGGGAATACGGGC	2248
Qy	2041	AGCAGTGAACCTGTGGAAGGGTTACAAAAGAGCTGTCTCATGTCATCTCTGATGAC	2100
Db	2249	AGCAGTGAACCTGTGGAAGGGTTACAAAAGAGCTGTCTCATGTCATCTCTGATGAC	2308
Qy	2101	AACCGGTTCAGCAAAATGATGGAATGCTCAAGCAGACAAAACCGCATGCTCTCTTTGTG	2160
Db	2309	AACCGGTTCAGCAAAATGATGGAATGCTCAAGCAGACAAAACCGCATGCTCTCTTTGTG	2368
Qy	2161	CAAGATGCACAGTCCCATGTGAGAGACCTGCACCTTCACCTGCTTGTCCAGTTTACG	2220
Db	2369	CAAGATGCACAGTCCCATGTGAGAGACCTGCACCTTCACCTGCTTGTCCAGTTTACG	2428
Qy	2221	CCCTGCTCCAGAACTGTGAAGCCACAAAAGTAGGCGCGACAGCTCACAGGGAAAAGC	2280
Db	2429	CCCTGCTCCAGAACTGTGAAGCCACAAAAGTAGGCGCGACAGCTCACAGGGAAAAGC	2488
Qy	2281	AGAAAGAGGAGAAATGCCAGGATTCGACCTTTACCTCTAGTGGAGCAGAACTATGT	2340
Db	2489	AGAAAGAGGAGAAATGCCAGGATTCGACCTTTACCTCTAGTGGAGCAGAACTATGT	2548
Qy	2341	CCTTGTGATGAATTTATATCCCAACCTTTATGGAACTGGTTCAGATTGCATTTCTTCAGAA	2400

MGC.  
 Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 6112)  
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
 Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,  
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,  
 Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
 Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
 Scheerz, T.E., Brownstein, M.J., Usdin, T.B., roshiyuki, S.,  
 Carninci, P., Prange, C., Raja, S.S., Loquellano, N.A., Peters, G.J.,  
 Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,  
 Worley, K.C., Valek, J.A., Gunaratne, P.H., Richards, S.,  
 Morley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
 Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
 Fahey, J., Helton, E., Kettner, M., Madan, A., Rodrigues, S.,  
 Sanchez, A., Whiting, M., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
 Bouffard, G.C., Rodriguez, A.C., Grimwood, J., Schmutz, J.,  
 Dackiw, M.G., Krzywicki, M.I., Skalska, U., Smalls, D.E.,  
 Butterfield, A., Schein, J.E., Jones, S.J., and Marra, M.A.  
 Genomewall.org  
 Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
 12477932  
 2 (bases 1 to 6112)  
 Strausberg, R.  
 Direct Submission  
 Submitted (02-DEC-2003) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 NIH-MGC Project URL: <http://mgc.ncbi.nih.gov>  
 Contact: MGC help desk  
 Email: [cgapbs@mail.nih.gov](mailto:cgapbs@mail.nih.gov)  
 Tissue Procurement: Dr. Jim Lin, University of Iowa  
 cDNA Library Preparation: The Bencio Soares, University of Iowa  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)  
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
 Center, Stanford University School of Medicine, Stanford, CA 94305  
 Web site: <http://www.shgc.stanford.edu>  
 Contact: (Dickson, Mark) [mcd@paxil.stanford.edu](mailto:mcd@paxil.stanford.edu)  
 (Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,  
 R. M.  
 Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LMNL at: <http://image.llnl.gov>  
 Series: IRAP Plate: 125 Row: i Column: 17  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 27369657  
 Location/Qualifiers  
 1. .6112  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone="MGC:66597 IMAGE:6410178"  
 /tissue\_type="Brain, enriched mouse brain 12.5dp"  
 /clone\_lib="NIH SWAP\_F00"  
 /lab\_host="DH10B"  
 /notes="Vector: pYX-ASC"  
 1. .6112  
 /gene="D130067I03Rik"  
 /note="synonym: B930082A18"  
 /db\_xref="LocusID:210417"  
 /db\_xref="MGSI:2443925"  
 162. .4985  
 /gene="D130067I03Rik"  
 /codon\_start=1  
 /product="n130067I03Rik protein"

GenCore version 5.1.6  
Copyright (C) 1993 - 2004 CompuGen Ltd.

cleic - nucleic search, using sw model  
n: February 23, 2004, 10:50:55 ; Search time 16521.2 Seconds  
(without alignments)  
11538.043 Million cell updates/sec  
US-10-022-710-1  
ct score: 4398  
nce: 1 atggtgcgtgcattcagaa.....atggagacttagacatgtaa 4398

ng table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0.  
hed: 3470272 segs, 21671516995 residues  
l number of hits satisfying chosen parameters: 6940544

num DB seq length: 0  
num DB seq length: 2000000000  
-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

case : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_ov.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_cm.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	4380.4	99.6	5724	9	AB051466
2	3321.2	75.5	6112	10	BC063250 Mus muscu
3	2200.2	50.0	2421	9	BC033125 Homo sapi
4	2200.2	50.0	2422	9	BC019344 Homo sapi
5	1232.8	28.0	5669	9	AB023177 Homo sapi
6	1232.8	28.0	6373	6	AX079870 Sequence
7	1229.6	28.0	6378	6	AX338841 Sequence
8	1129.6	25.7	5032	6	AX818221 Sequence
9	1109.8	25.2	5287	6	AX780247 Sequence
10	580.6	13.2	3053	6	AX747433 Sequence
11	580.6	13.2	3053	9	AK092252 Homo sapi
12	527	12.0	119205	9	AC011231 Homo sapi
13	527	12.0	167462	2	AC027600 Homo sapi
14	362.4	8.2	224789	2	AC111611 Rattus no
15	360.8	8.2	259236	2	AC123741 Mus muscu
16	270.4	6.1	169928	9	AC019067 Homo sapi
17	239.2	5.4	194472	9	AC012000 Homo sapi
18	198.4	4.5	225453	2	AC097957 Rattus no
19	196.4	4.5	85212	9	AC016679 Homo sapi
20	189.2	4.3	215703	2	AC121276 Mus muscu
21	175.8	4.0	68894	2	AC100201 Mus muscu
22	166.4	3.8	256545	2	AC110536 Mus muscu
23	165	3.8	259236	2	AC123741 Mus muscu
24	163.4	3.7	277376	2	AC128972 Rattus no
25	154	3.5	62063	2	AC136706 Mus muscu
26	154	3.5	183783	9	AC017082 Homo sapi
27	137.6	3.1	258545	2	AC110536 Mus muscu
28	133.8	3.0	399	11	G61408 SHGC-85947
29	132.2	3.0	198944	2	AC115912 Mus muscu
30	132	3.0	221262	2	AC122394 Mus muscu
31	128.8	2.9	872	11	BV023652 S212P6059
32	124.2	2.8	95921	9	AC004141 Homo sapi
33	124.2	2.8	168008	9	AC142334 Pan trogl
34	124	2.8	215510	2	AC128337 Rattus no
35	124	2.8	263244	2	AC097540 Rattus no
36	119.4	2.7	198944	2	AC115912 Mus muscu
37	117.6	2.7	110000	2	AC096315 Continuation (4 of
38	117.6	2.7	110000	2	AC096315 Continuation (5 of
39	116.8	2.7	256124	2	AC130447 Rattus no
40	108.8	2.5	191439	10	AL627302 Mouse DNA
41	108.8	2.5	195140	2	EX005251 Mus muscu
42	103.6	2.4	1704	9	BC016980 Homo sapi
43	102	2.3	300201	2	AC107086 Rattus no
44	100.2	2.3	143751	9	AC004160 Homo sapi
45	98.2	2.2	85693	2	AC124117 Mus muscu

ALIGNMENTS

RESULT 1  
AB051466  
LOCUS AB051466 5724 bp mRNA linear PRI 07-FEB-2001  
DEFINITION Homo sapiens mRNA for KIAA1679 protein, partial cds.  
ACCESSION AB051466  
VERSION AB051466.1 GI:12697902  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE  
AUTHORS Nagase, T., Kikuno, R., Hattori, A., Kondo, Y., Okumura, K. and Ohara, O.  
TITLE Prediction of the coding sequences of unidentified human genes.  
XIX. The complete sequences of 100 new cDNA clones from brain which



1469 GGACATCGTATTCTGAAGCCGCTCTCCCAAGATGACCGGGAAGAATATACGGGAGT 1528  
1321 CTTTGGCCAGTTCCTCCCTCTCTGAGAGAGAGTCTTGTGAATTCCTTCCGGAATGGAC 1380  
1529 CTTTGGCCAGTTCCTCCCTCTCTGAGAGAGAGTCTTGTGAATTCCTTCCGGAATGGAC 1588  
1381 TGTGTCTGAGAGAGTGGACGGAGTGTCTATCCCTGTTCCAGTCTCTGTTCAATAAATAAC 1440  
1589 TGTGTCTGAGAGAGTGGACGGAGTGTCTATCCCTGTTCCAGTCTCTGTTCAATAAATAAC 1648  
1441 TCAGATGGGAAAAGACAGCCAGTCAAGAACTATCTGGCACTGCTGGGGAAGTGGAAAG 1500  
1649 TCAGATGGGAAAAGACAGCCAGTCAAGAACTATCTGGCACTGCTGGGGAAGTGGAAAG 1708  
1501 CCATGTCCCCCTAGTCAAGGCTCTCCCAAGAGCATGTTTGTATGACCAATTCCTGTATG 1560  
1709 CCATGTCCCCCTAGTCAAGGCTCTCCCAAGAGCATGTTTGTATGACCAATTCCTGTATG 1768  
1561 CAGCTTCACTGGGAGACATCGCTTGGGCCCCCTGTTCTGAGGACACATTTGTAATGCC 1620  
1769 CAGCTTCACTGGGAGACATCGCTTGGGCCCCCTGTTCTGAGGACACATTTGTAATGCC 1828  
1621 CTTAATGCAACCAATGCTGAATGGGAAGCCACGTTGTGTGAGGATTCAGACTCGG 1680  
1829 CTTAATGCAACCAATGCTGAATGGGAAGCCACGTTGTGTGAGGATTCAGACTCGG 1888  
1681 AGAGTCTTCTGTCAGAGTCACTGGGAGCAAGTATGACCAAAAGATGTCAGATTCT 1740  
1889 AGAGTCTTCTGTCAGAGTCACTGGGAGCAAGTATGACCAAAAGATGTCAGATTCT 1948  
1741 ACTGACACTGAACTGTGGCCCCCTGTTTCTCCATGCAAAAAGACTGTATTGTGACT 1800  
1949 ACTGACACTGAACTGTGGCCCCCTGTTTCTCCATGCAAAAAGACTGTATTGTGACT 2008  
1801 GCTTTCAGTGTGAGACACCTTGGCCCAAGATGTGCCAAGCAGGAATGCCACAGTAAAA 1860  
2009 GCTTTCAGTGTGAGACACCTTGGCCCAAGATGTGCCAAGCAGGAATGCCACAGTAAAA 2068  
1861 CAGTCTCGATACAGAAATCATCATCCAGAGCAGCAATGGAGGCCAGGAATGCCAGAT 2128  
2069 CAGTCTCGATACAGAAATCATCATCCAGAGCAGCAATGGAGGCCAGGAATGCCAGAT 2180  
1921 ACCTTATATGAGAGAGAGTGTGAAGTTCCTGTTGTCTCTGATATCGGTGGAG 1980  
2129 ACCTTATATGAGAGAGAGTGTGAAGTTCCTGTTGTCTCTGATATCGGTGGAG 2188  
1981 CCACAGAAATGGAGCCCTTGCATCTTAGTGCCAGAGTCTGTCTGGCAGGGAATAACGGGC 2040  
2189 CCACAGAAATGGAGCCCTTGCATCTTAGTGCCAGAGTCTGTCTGGCAGGGAATAACGGGC 2248  
2041 AGCAGTGAAGCTGTGGAAAGGGGTTACAAAAGAGTGTCTCATGATCTCTGATGAC 2100  
2249 AGCAGTGAAGCTGTGGAAAGGGGTTACAAAAGAGTGTCTCATGATCTCTGATGAC 2308  
2101 ACCGGTCAAGCAATGATGGAATGCTCAAGCAGACAAACGGCATGCTCTCTTGTG 2160  
2309 ACCGGTCAAGCAATGATGGAATGCTCAAGCAGACAAACGGCATGCTCTCTTGTG 2368  
2161 CAAGAATGCACTGCCATGTCAGAGAGTGCACCTTCACTGCTTGGTCCAGTTTACG 2220  
2369 CAAGAATGCACTGCCATGTCAGAGAGTGCACCTTCACTGCTTGGTCCAGTTTACG 2428  
2221 CCCTGCTCCAGGACTGTGAAGCCACAAAGTAGGCGGCGGAGCTCAAGGGAAGAGC 2280  
2429 CCCTGCTCCAGGACTGTGAAGCCACAAAGTAGGCGGCGGAGCTCAAGGGAAGAGC 2488  
2281 AGAAGAAGAGGAATGCCAGGATTCCTGACCTTTACCTTTAGTGGAGACAGAACTATGT 2340  
2489 AGAAGAAGAGGAATGCCAGGATTCCTGACCTTTACCTTTAGTGGAGACAGAACTATGT 2548  
2341 CCTGTGTATGATTTATATCCACCTTATGGAACCTGGTCAAGATTGCTTCTCCAGAA 2400

2549 CCTTGTGATGAATTTATATATCCCAACCTTATGGAACCTGGTCAAGATTGCATTTCTCCAGAA 2608  
2401 GGCAGAGGAGGAGCTTCAACGAGGAGCTGGGGTACAAGACAGACAGCAAGAAATGTGGAGAA 2460  
2609 GGCAGAGGAGGAGCTTCAACGAGGAGCTGGGGTACAAGACAGACAGCAAGAAATGTGGAGAA 2668  
2461 GGCCTGGCGCTTTTGAGCAGTAGCTGTTCTGATAAATAATGGAAGACCTGTTGACCCCTCC 2520  
2669 GGCCTGGCGCTTTTGAGCAGTAGCTGTTCTGATAAATAATGGAAGACCTGTTGACCCCTCC 2728  
2521 TTTGCGACAGCTTCTGTTTACATCAAGAAATATGTCATTCCTGCCCATTTGATTGC 2580  
2729 TTTGCGACAGCTTCTGTTTACATCAAGAAATATGTCATTCCTGCCCATTTGATTGC 2788  
2581 AAGTTAAGCGATTGCTTCTAGTTGGGGCTTTGCGATTCATCTTGTGGAATTTGGAGTGAGA 2640  
2789 AAGTTAAGCGATTGCTTCTAGTTGGGGCTTTGCGATTCATCTTGTGGAATTTGGAGTGAGA 2848  
2641 ATTGATCCAAATGGCTTAAGAAAGAAACCTTACAAATGGAGGACGACCATGTCCTCAAACTG 2700  
2849 ATTGATCCAAATGGCTTAAGAAAGAAACCTTACAAATGGAGGACGACCATGTCCTCAAACTG 2908  
2701 GATCTCAAGAA-----TACGGTACATGAGGAGCTCCCATGTTTACAGTGAAGTCAATCAG 2754  
2909 GATCTCAAGAAATCAGGCTCAAGTACATGAGGAGCTCCCATGTTTACAGTGAAGTCAATCAG 2968  
2755 TATTCTCGGTTGTAGAACACTGTTCTTATGCAAAATCAACAAATGAGCTGAGTGCCTG 2814  
2969 TATTCTCGGTTGTAGAACACTGTTCTTATGCAAAATCAACAAATGAGCTGAGTGCCTG 3028  
2815 CCGTGTGGAGGAGGAACAACATCTAGGAAATTCAGATGTGAAATCTCGGATGGTGA 2874  
3029 CCGTGTGGAGGAGGAACAACATCTAGGAAATTCAGATGTGAAATCTCGGATGGTGA 3088  
2875 GGTGGAGCAGTGCATAGCAACCTGTGCAACCAAGATGAATTCCTCCAGAAACCCAGTCC 2934  
3089 GGTGGAGCAGTGCATAGCAACCTGTGCAACCAAGATGAATTCCTCCAGAAACCCAGTCC 3148  
2935 TGTTCCTTATGTTGCTCCCAATCAGTGTGTCATGTTCTGAGTGGGGACTTTGGAGCAAAATGC 2994  
3149 TGTTCCTTATGTTGCTCCCAATGAGTGTGTCATGTTCTGAGTGGGGACTTTGGAGCAAAATGC 3208  
2995 CCACAGTCTATGGATCCCAACATGAGAGAGAACTGCGCACTGCTAGAGCAATCA 3054  
3209 CCACAGTCTATGGATCCCAACATGAGAGAGAACTGCGCACTGCTAGAGCAATCA 3268  
3055 CTGAATCTCAGGACTTGTGCTGAAGACTCAGAGTGCAGCTTGTCTCTGATGAATGAAAT 3114  
3269 CTGAATCTCAGGACTTGTGCTGAAGACTCAGAGTGCAGCTTGTCTCTGATGAATGAAAT 3328  
3115 TGTCTCCAGTTCCAGTACAAATCTAAACAGAGTGGAGCACTGCGAGCTGAGTGAACAGCA 3174  
3329 TGTCTCCAGTTCCAGTACAAATCTAAACAGAGTGGAGCACTGCGAGCTGAGTGAACAGCA 3388  
3175 CCCTGTGGTCAAGGCTCAGGACCCCTGCTAAGCTGTGTGTCAGTGTATGGCAAGCA 3234  
3389 CCCTGTGGTCAAGGCTCAGGACCCCTGCTAAGCTGTGTGTCAGTGTATGGCAAGCA 3448  
3235 GTCAAGATGGAACCAATGTGAGCAGCATTAATTTGGAGAGCCCAAGAGAAATGAGCATTC 3294  
3449 GTCAAGATGGAACCAATGTGAGCAGCATTAATTTGGAGAGCCCAAGAGAAATGAGCATTC 3508  
3295 TGTCTGTGGAATGCGTGTCACTGTGCTGAGCTCTCAGGCTGGAGCGGCTTGGACAGAGTGT 3354  
3509 TGTCTGTGGAATGCGTGTCACTGTGCTGAGCTCTCAGGCTGGAGCGGCTTGGACAGAGTGT 3568  
3355 TCACAGACCTGTGGCCCATGGAGTGCAGTGTGAGTGCAGTGTGAGTGCAGTGTGAGTGC 3414  
3569 TCACAGACCTGTGGCCCATGGAGTGCAGTGTGAGTGCAGTGTGAGTGCAGTGTGAGTGC 3628  
3415 CAAGAGAGAGAGCGGCAATGCCCCACAGAGCTTACCCAGGAGAAACCTGCCAGTGCAC 3474  
3629 CAAGAGAGAGAGCGGCAATGCCCCACAGAGCTTACCCAGGAGAAACCTGCCAGTGCAC 3688

```

3475 CCCTGCTACAGCTGGTCTCTGGCACTGGTCTGATGTAATTTGGAGGTGGAGACTGT 3534
|||||
3689 CCCTGCTACAGCTGGTCTCTGGCACTGGTCTGATGTAATTTGGAGGTGGAGACTGT 3748
|||||
3535 GGGGAAGAGAGTTTCAGATCCGACGCTTTCTGATCGTGTCCAGTGGTTCAATATCTCAT 3594
|||||
3749 GGGGAAGAGAGTTTCAGATCCGACGCTTTCTGATCGTGTCCAGTGGTTCAATATCTCAT 3808
|||||
3595 GCAGCTGGAGCTGTGAGAGATGCACTGTGTGAGAAATGCCCTTTTCAGAGACGATCTCG 3654
|||||
3809 GCAGCTGGAGCTGTGAGAGATGCACTGTGTGAGAAATGCCCTTTTCAGAGACGATCTCG 3868
|||||
3655 AAGCAGCTGTGTCTGCTGCTTCCAGAGAGCTGCCATTTAACAGATGGTTCAGAGTGG 3714
|||||
3869 AAGCAGCTGTGTCTGCTGCTTCCAGAGAGCTGCCATTTAACAGATGGTTCAGAGTGG 3928
|||||
3715 AGCAGATGTAATTAACCTGCAATGTGATGGAAGAGCTTTGAGACTGTGGCGCCGACGTCT 3774
|||||
3929 AGCAGATGTAATTAACCTGCAATGTGATGGAAGAGCTTTGAGACTGTGGCGCCGACGTCT 3988
|||||
3775 AGATCAAGAGCTTTTATAATTCAGTCTTTTGAGAACCAAGACAGCTGCCCCCAACAGTGT 3834
|||||
3989 AGATCAAGAGCTTTTATAATTCAGTCTTTTGAGAACCAAGACAGCTGCCCCCAACAGTGT 4048
|||||
3835 CTAGAAACACGCGCTTTGTACAGAGGCAATGTTATCACTACACATGGAAGCAAGTCTT 3894
|||||
4049 CTAGAAACACGCGCTTTGTACAGAGGCAATGTTATCACTACACATGGAAGCAAGTCTT 4108
|||||
3895 TCGAACAATTAACGAACTGATGCTGCTGCCAGCTTCAGATGGCGTTAATGTCACAGA 3954
|||||
4109 TCGAACAATTAACGAACTGATGCTGCTGCCAGCTTCAGATGGCGTTAATGTCACAGA 4168
|||||
3955 GGCTGCTCCCTCAGGCGCGTCTGCTGCCATTCGGCAGTGCATTCACGCTTCGCAAAA 4014
|||||
4169 GGCTGCTCCCTCAGGCGCGTCTGCTGCCATTCGGCAGTGCATTCACGCTTCGCAAAA 4228
|||||
4015 CCTTCTCTCTACTGTACACAGGTGGAGTCTGCTGTGTGAGAGGGCTATACAGAGATA 4074
|||||
4229 CCTTCTCTCTACTGTACACAGGTGGAGTCTGCTGTGTGAGAGGGCTATACAGAGATA 4288
|||||
4075 ATGAATCAAAATGGTTCTCTGATTAATCTGATGAAAGTACCAAGGCTCAGAGATAAAAA 4134
|||||
4289 ATGAGATCAAAATGGTTCTCTGATTAATCTGATGAAAGTACCAAGGCTCAGAGATAAAAA 4348
|||||
4135 GCTGATGTGAAACCTTTCTGGGAAAAACAGACCTGTGAAATTCAGAAATATCATGATTT 4194
|||||
4349 GCTGATGTGAAACCTTTCTGGGAAAAACAGACCTGTGAAATTCAGAAATATCATGATTT 4408
|||||
4195 TTTAAAGGATGGTCTCTTCAACCACTTGATCCAGATGGCGGATGAAATTTGGGTTTAT 4254
|||||
4409 TTTAAAGGATGGTCTCTTCAACCACTTGATCCAGATGGCGGATGAAATTTGGGTTTAT 4468
|||||
4255 GCGTTTTCAGTGGCGCTTTCTCATCATGATTTTCTTAATATTTTACTTCTTACCTTGT 4314
|||||
4469 GCGTTTTCAGTGGCGCTTTCTCATCATGATTTTCTTAATATTTTACTTCTTACCTTGT 4528
|||||
4315 TGCAGAGAGCCAAACACATCAAGACACCTCCCAACAGAGGCTCTGACCTTAGCC 4374
|||||
4529 TGCAGAGAGCCAAACACATCAAGACACCTCCCAACAGAGGCTCTGACCTTAGCC 4588
|||||
4375 TAGCATGGAGACTTAGACATGTAA 4398
|||||
4589 TAGCATGGAGACTTAGACATGTAA 4612
|||||

KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 6112)
AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, P.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, J., Max, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Udén, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullaby, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Morley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalek, U., Smailus, D.E.,
Schnerich, A., Schein, J.E., Jones, S.J., and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 6112)
Strausberg, R.
Direct Submission
Submitted (02-DEC-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NTH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: gcgaps-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library Preparation: M. Bento Soares, University of Iowa
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
cDNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAX Plate: 125 Row: i Column: 17
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 27369657.
Location/Qualifiers
1. 6112
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="MGC:66597 IMAGE:6410178"
/tissue_type="Brain, enriched mouse brain 12.5dp"
/clone_lib="NIH BMAP_F00"
/lab_host="DH10B"
/vec="Vector: pYX-ASC"
1. 6112
/gene="D130067I03Rik"
/notes="synonym: B930082A18"
/db_xref="LocusID:210417"
/db_xref="MGI:2443925"
162. 4985
/gene="D130067I03Rik"
/codon_start=1
/product="D130067I03Rik protein"

FEATURES
source
gene
CDS
BC063250 6112 bp mRNA linear ROD 16-DEC-2003
N Mus musculus RIKEN cDNA D130067I03 gene, mRNA (cDNA clone MGC:66597
IMAGE:6410178), complete cds.
BC063250
BC063250.1 GI:38649124

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

DM nucleic - nucleic search, using sw model  
Run on: February 23, 2004, 10:50:55 ; Search time 16521.2 Seconds  
(without alignments)  
11538.043 Million cell updates/sec

Title: US-10-022-710-1  
Perfect score: 4398  
Sequence: 1 atgggtcgtgcattcagaa.....atggagacttagacatgtaa 4398

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues  
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_cm.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_or.\*
- 21: em\_ov.\*
- 22: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rtd.\*
- 36: em\_htg\_rtd.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4380.4	99.6	5724	9	AB051466	AB051466 Homo sapi
2	3321.2	75.5	6112	10	BC063250	BC063250 Mus muscu
3	2200.2	50.0	2421	9	BC033125	BC033125 Homo sapi
4	2200.2	50.0	2422	9	BC019344	BC019344 Homo sapi
5	1232.8	28.0	5669	9	AB023177	AB023177 Homo sapi
6	1232.8	28.0	6373	6	AX079870	AX079870 Sequence
7	1229.6	28.0	6378	6	AX079870	AX079870 Sequence
8	1129.6	25.7	5032	6	AX818221	AX818221 Sequence
9	1109.8	25.2	5287	6	AX780247	AX780247 Sequence
10	580.6	13.2	3053	6	AX747433	AX747433 Sequence
11	580.6	13.2	3053	9	AX092252	AX092252 Homo sapi
12	527	12.0	119205	9	AC011231	AC011231 Homo sapi
13	527	12.0	167462	2	AC027600	AC027600 Homo sapi
14	362.4	8.2	224789	2	AC111611	AC111611 Rattus no
15	360.8	8.2	259236	2	AC123741	AC123741 Mus muscu
16	270.4	6.1	169928	9	AC019067	AC019067 Homo sapi
17	239.2	5.4	194472	9	AC012000	AC012000 Homo sapi
18	198.4	4.5	225453	2	AC097957	AC097957 Rattus no
19	196.4	4.5	89212	9	AC016679	AC016679 Homo sapi
20	189.2	4.3	215703	2	AC121276	AC121276 Mus muscu
21	175.8	4.0	68894	2	AC100201	AC100201 Mus muscu
22	166.4	3.8	256545	2	AC110536	AC110536 Mus muscu
23	165	3.8	259236	2	AC123741	AC123741 Mus muscu
24	163.4	3.7	277376	2	AC128972	AC128972 Rattus no
25	154	3.5	62063	2	AC136706	AC136706 Mus muscu
26	154	3.5	183783	9	AC017082	AC017082 Homo sapi
27	137.6	3.1	256545	2	AC110536	AC110536 Mus muscu
28	133.8	3.0	399	11	G61408	G61408 SHGC-85947
29	132.2	3.0	198944	2	AC115912	AC115912 Mus muscu
30	132	3.0	221262	2	AC123394	AC123394 Mus muscu
31	128.8	2.9	872	11	BV023652	BV023652 S212P6059
32	124.2	2.8	95921	9	AC004141	AC004141 Homo sapi
33	124.2	2.8	168008	9	AC142334	AC142334 Pan trogl
34	124	2.8	215510	2	AC128337	AC128337 Rattus no
35	119.4	2.7	198944	2	AC097957	AC097957 Rattus no
36	117.6	2.7	110000	2	AC096315	AC096315 Mus muscu
37	117.6	2.7	110000	2	AC096315	Continuation (4 of
38	117.6	2.7	110000	2	AC096315	Continuation (5 of
39	116.8	2.7	256124	2	AC130447	AC130447 Rattus no
40	108.8	2.5	191439	10	AL627302	AL627302 Mouse DNA
41	108.8	2.5	191440	2	EX005251	EX005251 Mus muscu
42	103.6	2.4	1704	9	BC016980	BC016980 Homo sapi
43	102	2.3	300201	2	AC107086	AC107086 Rattus no
44	100.2	2.3	143751	9	AC004160	AC004160 Homo sapi
45	98.2	2.2	85693	2	AC124117	AC124117 Mus muscu

ALIGNMENTS

RESULT 1  
AB051466  
LOCUS AB051466 5724 bp mRNA linear PRI 07-FEB-2001  
DEFINITION Homo sapiens mRNA for KIAA1679 protein, partial cds.  
ACCESSION AB051466  
VERSION AB051466.1 GI:12697902  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (sites)  
AUTHORS Nagase,T., Kikuno,R., Hattori,A., Kondo,Y., Okumura,K. and Ohara,O.  
TITLE Prediction of the coding sequences of unidentified human genes.  
XIX. The complete sequences of 100 new cDNA clones from brain which



code for large proteins in vitro  
DNA Res. 7 (6), 347-355 (2000).  
21082932  
11214970  
2 (bases 1 to 5724)  
Chara, O., Nagase, T. and Kikuno, R.  
Direct Submission  
Submitted (22-NOV-2000) Osemu Ohara, Kazusa DNA Research Institute,  
Department of Human Gene Research; 1532-3, Yana, Kisarazu, Chiba  
292-0812, Japan (E-mail: cdnainfo@kazusa.or.jp)  
URL: http://www.kazusa.or.jp/huge, Tel: 81-438-52-3913,  
Fax: 81-438-52-3914

Location/Qualifiers  
1..5724  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="fh22954"  
/note="vector:pBluescriptII SK plus"  
1..5724  
/gene="XIAA1679"  
1..4612  
/gene="XIAA1679"  
/note="Start codon is not identified."

/codon\_start=2  
/product="XIAA1679 protein"  
/protein\_id="BAB21770.1"  
/db\_xref="GI:12697903"  
/translation="GWTGHLNCGESNRPKRSQFRVCDWHSDLFOWESDMHHCVL  
VPARGVEKPTAEVCVTAQHGLQHRMVCIOKLNRTVVANEICBHFALOPTEOACLI  
PCPRCIVSEBPLPWSNCKSGGKGLQHTRAVIAPLFGGLQCNLPESBACDAPISC  
PLGESEYFSLKVGWPKSKRPLHLKBNPSGRTVLDFNSDSNERVTPKHQSYKAHHHS  
KSWALEIGYQTRQVSTSDGONAMLSLCLQDSFPLTVQSCIMPEKDCETOWMSWSPC  
SKTCSGLLQPTFRSRSNRKXMAIGGKCEPILLEKEACIEGELLEQQCPYRSVTR  
EWKEQCSILLLEQQPHVHTVPGVCGGIQTRVYCAQSVPAALAAKAREVREVEKA  
LCVGPALPISOLNCPSTDCITVSSWANGLCIENHCEHPOCKKFRTRORHLVEST  
GPAGCHPLVESVPCEDPMCVRWLASGEGICFPHGCKGLGHRILKAVCONDRGDSG  
SLCPVPPPERKSCIPCEMOCVLEWTEMSSCSQSKNSDKGKTSRSTILALAGE  
GKPCPPSQAOLGHRKLCNDSMQLHNETISFWGSCSEDTLVNALNIGWNGEATCGV  
GIQTRVFCVSHVGQVMTKCPDSTRPETVRPFLPKCKDCI VTAASEWTPCPMRCQ  
AGNATVKQSRVRIIOEAANGQCEPDLYEERECEDVSLCPVYFARSPKQWSPCLVP  
ESVMQITGSSBAGKIQTRAVSCISDNRSAEMECIKCTNGMPLLVORCTVPCRE  
DCTFAWKTGTPCSTNCASTRQLTKRKEKCKODSLYPLVELTCLPCDEFIS  
OYGNWSICILPEGRREPHGLRVQADSKEGEGELFRFAVACDKNGRPVDPSPCSSS  
GIQKHEVIPCPEFDCLEDSWWSNGSCSSGIGVIRSKWLKEXKPYNGRRPCPLDLK  
NOAQHEAVPCYSECNQISWVHWSSCKINNELSLRCGSGGSRKIRCVNTADGEG  
GAVDSNLNQDEIIPETQSCSLMCPNECWSEWGLWPKQSCDPHTMQRTREHLRP  
SLNSRTCAEDSQVQCLINENCFOQYNLTEWSTQLESENAPCGQVTRLLSCVSD  
GKPVSMDOCEQHNLEKPOKMSIPCLVECVWQCLSGMTAWECSQTCGHGGRMSTRF  
IIMPTQEGRRPCPTLTQKTCPTVPCYSWLGNWSACKLEGGCGEGVQ:RSLSCMV  
HSGSLSHAGRVEDALCGMPQDSILKOLCSVPCPGDCHLTWSENSTCELTCDGR  
SFETVGRQSRSTFIIQSFENQDSQPQVLETRPCTGKCYHYTWKSLMNNERTVW  
CQSDGVNVTGSCSPQAPAAIRQCIIPACRRFPSTCTGGVCGCKEKTIELMNSGFL  
DYCMKVPGEEDKKADVKNLSGKRNVPNSKIHIDIFKWSLSQLDPLDPDPVKIWIYGVSGG  
AFILMIFLIFTSYLVCKPKPQSTPPQOKPLTLAYDGLDY"

## ORIGIN

Query Match 99.6%; Score 4380.4; DB 9; Length 5724;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 4397; Conservative 0; Mismatches 1; Indels 6; Gaps 1;  
1 ATGGTCGGTGCATTAGAGGCTGAAACCGAACTGTGGTGGTTCGAAATGAAATATGCAACAC 60  
209 ATGGTCGGTGCATTAGAGGCTGAAACCGAACTGTGGTGGTTCGAAATGAAATATGCAACAC 268  
61 TTGCGCCTTCAGCCTCTCAGAACAGGCTTGCTCATTCCTTGCTCCCGGGAATTGTGTA 120  
269 TTGCGCCTTCAGCCTCTCAGAACAGGCTTGCTCATTCCTTGCTCCCGGGAATTGTGTA 328  
121 GTATCTGAGTTCCTTACATGCTCCAACTGTAGCAAGGATGTGGGAAGAAATTCAGAGCAT 180  
329 GTATCTGAGTTCCTTACATGCTCCAACTGTAGCAAGGATGTGGGAAGAAATTCAGAGCAT 388

181 AGAACTCGCGCGGTGCATAGCTCCCTCTCTTTGGTGGTTCGAAATGCAATCTGACT 240  
389 AGAACTCGCGCGGTGCATAGCTCCCTCTCTTTGGTGGTTCGAAATGCAATCTGACT 448  
241 GAGTCAAGAGCCTGTGATGCTCCCATTTCTCTCTTGGGAGAGGAAATATACATTT 300  
449 GAGTCAAGAGCCTGTGATGCTCCCATTTCTCTCTTGGGAGAGGAAATATACATTT 508  
301 AGCCTTAAGGTTGGACCATGAGTAAATGACAGCTGCCTCATCTTTAAAGAAATTAATCCA 360  
509 AGCCTTAAGGTTGGACCATGAGTAAATGACAGCTGCCTCATCTTTAAAGAAATTAATCCA 568  
361 AGCGGAAGAACTGTGCTGATTTTAACTCTGATTTAAATGACGAGTCACTTTAAACAT 420  
569 AGCGGAAGAACTGTGCTGATTTTAACTCTGATTTAAATGACGAGTCACTTTAAACAT 628  
421 CAAAGTTACAAAGCACATCATCATTCGAAAGTCTTGGGCAATAGAGATAGTTATCAAACC 480  
629 CAAAGTTACAAAGCACATCATCATTCGAAAGTCTTGGGCAATAGAGATAGTTATCAAACC 688  
481 CGGCAGGTTTCGTGTACAAGAGTGAATGACAAATGCTATGTTAAGCCTTTGCGCTTCAA 540  
689 CGGCAGGTTTCGTGTACAAGAGTGAATGACAAATGCTATGTTAAGCCTTTGCGCTTCAA 748  
541 GATTCTCTCCCATTTGACTTCTCAGTCTCATCATGCCCCAAAGACTGTGAAACCTCCACG 600  
749 GATTCTCTCCCATTTGACTTCTCAGTCTCATCATGCCCCAAAGACTGTGAAACCTCCACG 808  
601 TGCTCTCTCTCGAGCCCTGCTCCAGACATCCGTTTCAGGAGTCTCTTTCAGGAGTTT 660  
809 TGCTCTCTCTCGAGCCCTGCTCCAGACATCCGTTTCAGGAGTCTCTTTCAGGAGTTT 868  
661 AGGAGCAGAGCCGGAACCTGGAAGCATGCTGATTTGGAGTGGGAAAGAGTCTCTGAA 720  
869 AGGAGCAGAGCCGGAACCTGGAAGCATGCTGATTTGGAGTGGGAAAGAGTCTCTGAA 928  
721 CTCTCTTGAAGAAAGAGCCCTGCATTTGTTGAAGAGAACTTCTTCAGCAATGTCCTCCAGTAT 780  
929 CTCTCTTGAAGAAAGAGCCCTGCATTTGTTGAAGAGAACTTCTTCAGCAATGTCCTCCAGTAT 988  
781 TCCTGAGAACTTCTGAATGGAAGAAATGCAAGTCTCTCTCTCTCGAGCAGAGGAT 840  
989 TCCTGAGAACTTCTGAATGGAAGAAATGCAAGTCTCTCTCTCTCGAGCAGAGGAT 1048  
841 CCCACTGCGCATGTGAGGAGCCCGTGTGTGGCGTGGGATCCAGACCCGAGAGTGTAC 900  
1049 CCCACTGCGCATGTGAGGAGCCCGTGTGTGGCGTGGGATCCAGACCCGAGAGTGTAC 1108  
901 TGTGCCCAGAGCGTACAGCAGCTGCCGCACTGAGGGCCAAAGAACTCTCTAGACCTGTG 960  
1109 TGTGCCCAGAGCGTACAGCAGCTGCCGCACTGAGGGCCAAAGAACTCTCTAGACCTGTG 1168  
961 GAAAAGGCATATGCTGGAGCCCGCCCTGTCCTCTCAGCTCTGCAATATCCCTTGC 1020  
1169 GAAAAGGCATATGCTGGAGCCCGCCCTGTCCTCTCAGCTCTGCAATATCCCTTGC 1228  
1021 TCTACGAGCTGCATAGTATCTCTCTGCTCAGCTCGGGGCTGTGCAATCCATGAAACTGT 1080  
1229 TCTACGAGCTGCATAGTATCTCTCTGCTCAGCTCGGGGCTGTGCAATCCATGAAACTGT 1288  
1081 CATGAACCTCAGGGGAAAAGAAAGATTAGAACAGGAGCGCCATCTCTCATGGAATCT 1140  
1289 CATGAACCTCAGGGGAAAAGAAAGATTAGAACAGGAGCGCCATCTCTCATGGAATCT 1348  
1141 ACAGGGCTCTCAGGGCAATTCCTCTCAATTTGGTGGAGTCTGTTCTTGTGAGGATCCAAATG 1200  
1349 ACAGGGCTCTCAGGGCAATTCCTCTCAATTTGGTGGAGTCTGTTCTTGTGAGGATCCAAATG 1408  
1201 TGCTACCGATGGCTGGCATCAGAAAGGATCTGTTTCCCTGATCATGGAATAATGSCCTG 1260  
1409 TGCTACCGATGGCTGGCATCAGAAAGGATCTGTTTCCCTGATCATGGAATAATGSCCTG 1468  
1261 GGACATCGTATCTGAGGCGCGTCTGCCAATGACCGCGGAGAGATGTATCAGGGAGT 1320

Db 1469 GGACATCGTATTCTGAAGGCGGTCTGCCAGATGACCGCGAGAGATGATCAGGAGT 1528  
QY 1321 CTTTGGCCAGTCCCTCTCTCTGAGAGAGTCTTGTGAATTCCTGCCAATGGAC 1380  
Db 1529 CTTTGGCCAGTCCCTCTCTCTGAGAGAGTCTTGTGAATTCCTGCCAATGGAC 1588  
QY 1381 TGTGTGCTGAGCGAGTGGACGGAGTGTCTATCCTGTTCCTGAGTCCAGTCTGTTCCTGTTCAATAAAAAAC 1440  
Db 1589 TGTGTGCTGAGCGAGTGGACGGAGTGTCTATCCTGTTCCTGAGTCCAGTCTGTTCCTGTTCAATAAAAAAC 1648  
QY 1441 TCAGATGGGAAAACAGACAGAGTCAAGAACTATCTGSCACTGGCTGGGAGAGTGGAAAG 1500  
Db 1649 TCAGATGGGAAAACAGACAGAGTCAAGAACTATCTGSCACTGGCTGGGAGAGTGGAAAG 1708  
QY 1501 CCATGTCCCTCTAGTCAAGGCTCTCCAGAGACATGTTTGTGTATGACCAATTCCTGTATG 1560  
Db 1709 CCATGTCCCTCTAGTCAAGGCTCTCCAGAGACATGTTTGTGTATGACCAATTCCTGTATG 1768  
QY 1561 CAGCTTCACTGGGAGACATGCTTGGGGCCCTTGTCTGAGGACACATGTTGTAATGCTGCC 1620  
Db 1769 CAGCTTCACTGGGAGACATGCTTGGGGCCCTTGTCTGAGGACACATGTTGTAATGCTGCC 1828  
QY 1621 CTTAATGCAACATTTGCTGGAATGGAGAGACCAAGTGTGTGTAGGCAATTCAGACTCGG 1680  
Db 1829 CTTAATGCAACATTTGCTGGAATGGAGAGACCAAGTGTGTGTAGGCAATTCAGACTCGG 1888  
QY 1681 AGAGTCTTCTGTGTCAGAGATCAGTGGGACAGTAAATGACCAAAAGATGTCAGACTTCT 1740  
Db 1889 AGAGTCTTCTGTGTCAGAGATCAGTGGGACAGTAAATGACCAAAAGATGTCAGACTTCT 1948  
QY 1741 ACTCGACTGAACTGTGGCCCTGTTTCTCCATGSCAAAAGACTCTATTGTGACT 1800  
Db 1949 ACTCGACTGAACTGTGGCCCTGTTTCTCCATGSCAAAAGACTCTATTGTGACT 2008  
QY 1801 GCTTTCAGTGAAGTGGACACCTGCCAAGGATGTGCCAAGCAGGAAATGCCAGTAAAA 1860  
Db 2009 GCTTTCAGTGAAGTGGACACCTGCCAAGGATGTGCCAAGCAGGAAATGCCAGTAAAA 2068  
QY 1861 CAGTCTCGATACAGAAATCATCCAGAGACAGCCAAATGGAGCCAGGAATGCCAGAT 1920  
Db 2069 CAGTCTCGATACAGAAATCATCCAGAGACAGCCAAATGGAGCCAGGAATGCCAGAT 2128  
QY 1921 ACCTTATATGAGGAGAGAGTGTGAAGATGTTTCTTGTGTCTGTATATCGTGGAG 1980  
Db 2129 ACCTTATATGAGGAGAGAGTGTGAAGATGTTTCTTGTGTCTGTATATCGTGGAG 2188  
QY 1981 CCACAGAAATGGAGCCCTTGATCTTATGTGCGAGTCTGTGTGGAGGAAATAACGGGC 2040  
Db 2189 CCACAGAAATGGAGCCCTTGATCTTATGTGCGAGTCTGTGTGGAGGAAATAACGGGC 2248  
QY 2041 AGCAGTGAAGCTGTGGAAGGGTTACAAACAGAGCTGTCTATGCATCTCTGATGAC 2100  
Db 2249 AGCAGTGAAGCTGTGGAAGGGTTACAAACAGAGCTGTCTATGCATCTCTGATGAC 2308  
QY 2101 AACCGGTTCAGCAGAAATGATGAAATGCTCAAGCAGACAAACGGCATGCCCTCTCTTGTG 2160  
Db 2309 AACCGGTTCAGCAGAAATGATGAAATGCTCAAGCAGACAAACGGCATGCCCTCTCTTGTG 2368  
QY 2161 CAAGNATGCAAGTCCCATGTGAGAGACTGTGACCTTCACTGTTGTTGTTCAAGTTAAG 2220  
Db 2369 CAAGNATGCAAGTCCCATGTGAGAGACTGTGACCTTCACTGTTGTTGTTCAAGTTAAG 2428  
QY 2221 CCCTGCTCCAGAACTGTGAAGCCACAAAAGTAGTGGGGGGGACAGCTCAACAGGAAAAAGC 2280  
Db 2429 CCCTGCTCCAGAACTGTGAAGCCACAAAAGTAGTGGGGGGGACAGCTCAACAGGAAAAAGC 2488  
QY 2281 AGAAGAGAGGAAATGCCAGATTCGACCTTTACCTCTAGTGGAGACAGAACTATGT 2340  
Db 2489 AGAAGAGAGGAAATGCCAGATTCGACCTTTACCTCTAGTGGAGACAGAACTATGT 2548  
QY 2341 CCTTGTGATGAATTTATATCCCACTTATGAAACTGGTCAGATTCGATTCCTCCAGAA 2400

Db 2549 CCTTGTGATGAATTTATATCCCAACCTTATGAAAACCTGGTCAGATTGCAATTCCTCCAGAA 2608  
QY 2401 GCGAAGAGGAGCTCACCGAGAGCTCGGGTACAAGCAGACAGACAAAGAAATGTGGAGAA 2460  
Db 2609 GCGAAGAGGAGCTCACCGAGAGCTCGGGTACAAGCAGACAGACAAAGAAATGTGGAGAA 2668  
QY 2461 GGCCTGCGCTTTCGAGCAGTATGCTTCTGATAAAAATGGAAGACCTGTTGAGCCCTCC 2520  
Db 2669 GGCCTGCGCTTTCGAGCAGTATGCTTCTGATAAAAATGGAAGACCTGTTGAGCCCTCC 2728  
QY 2521 TTCTGACAGCTCTGGTTACATTCAGAAAAATGTGTCAATTCCTGCTGCCATTTGATTGC 2580  
Db 2729 TTCTGACAGCTCTGGTTACATTCAGAAAAATGTGTCAATTCCTGCTGCCATTTGATTGC 2788  
QY 2581 AAGTTAAAGCAATGGTCTAGTGGGGTCTTTCAGATTCATCTTGTGGAATTTGGAGTGAGA 2640  
Db 2789 AAGTTAAAGCAATGGTCTAGTGGGGTCTTTCAGATTCATCTTGTGGAATTTGGAGTGAGA 2848  
QY 2641 ATTCCATCCAAATGGCTAAAAAGAAAACCTTACATGGAGGACGACCATGTCCCAAACTG 2700  
Db 2849 ATTCCATCCAAATGGCTAAAAAGAAAACCTTACATGGAGGACGACCATGTCCCAAACTG 2908  
QY 2701 GATCTCAAGAA-----TGAAGTACATGAGGACGTCCCATGTTTACAGTGAAGTCAATCAG 2754  
Db 2909 GATCTCAAGAAATCAGGCTCAGTACATGAGGACGTCCCATGTTTACAGTGAAGTCAATCAG 2968  
QY 2755 TATTCCTGCTGTGAGACACTGGTCTTCATGCAAAATCAACATGAGTGAAGTCCCTG 2814  
Db 2969 TATTCCTGCTGTGAGACACTGGTCTTCATGCAAAATCAACATGAGTGAAGTCCCTG 3028  
QY 2815 CCGCTGAGAGAGGAAACAAATCTAGGAAAATCAGATGTGTGAATCTGCGGATGGTGAA 2874  
Db 3029 CCGCTGAGAGAGGAAACAAATCTAGGAAAATCAGATGTGTGAATCTGCGGATGGTGAA 3088  
QY 2875 GGTGAGCAGTGGATAGCAACCTGTGCAACAGAGATGAAATTCCTCCAGAAAACCCAGTCC 2934  
Db 3089 GGTGAGCAGTGGATAGCAACCTGTGCAACAGAGATGAAATTCCTCCAGAAAACCCAGTCC 3148  
QY 2935 TGTCTCTTATGTGTCCCAATGAGTGTGTCTGAGTGGGACTTTGGAGCAAAATGC 2994  
Db 3149 TGTCTCTTATGTGTCCCAATGAGTGTGTCTGAGTGGGACTTTGGAGCAAAATGC 3208  
QY 2995 CCACAGTCAATGCGATCCCCACACAAATCAGAGAGAAATCTGCGCACTCTGAAGCAATCA 3054  
Db 3209 CCACAGTCAATGCGATCCCCACACAAATCAGAGAGAAATCTGCGCACTCTGAAGCAATCA 3268  
QY 3055 CTGAATCAGAGACTGTGCTGAAGACTCAGAGTGCAGCCTTGCTCTCTGAATGAAAT 3114  
Db 3269 CTGAATCAGAGACTGTGCTGAAGACTCAGAGTGCAGCCTTGCTCTCTGAATGAAAT 3328  
QY 3115 TGTCTCCAGTTCAGTACAAATCTAACAGAGTGGAGCACATGSCAGCTGAGTGAAGAACGCA 3174  
Db 3329 TGTCTCCAGTTCAGTACAAATCTAACAGAGTGGAGCACATGSCAGCTGAGTGAAGAACGCA 3388  
QY 3175 CCGTGTGTCAGAGGCTGAGGACCCGCTGTAACTCAGAGTGCAGCCTTGCTCTCTGAAGCA 3234  
Db 3389 CCGTGTGTCAGAGGCTGAGGACCCGCTGTAACTCAGAGTGCAGCCTTGCTCTCTGAAGCA 3448  
QY 3235 GTCAGATGAGCAATGTGAGCAGCATTAATTTGGAGAGCCGCCAGAGAAATGAGCAATTC 3294  
Db 3449 GTCAGATGAGCAATGTGAGCAGCATTAATTTGGAGAGCCGCCAGAGAAATGAGCAATTC 3508  
QY 3295 TGTCTGTTGGAATGCGTGGTCAACTGTGTCAGCTCTCAGGGTGGACCGCTTGGACAGAGTGT 3354  
Db 3509 TGTCTGTTGGAATGCGTGGTCAACTGTGTCAGCTCTCAGGGTGGACCGCTTGGACAGAGTGT 3568  
QY 3355 TCACAGACCTGTGGCCATGGAGTCCGAATGAGCCGACTTCGATTTCATTTATGCAAC 3414  
Db 3569 TCACAGACCTGTGGCCATGGAGTCCGAATGAGCCGACTTCGATTTCATTTATGCAAC 3628  
QY 3415 CAAGGAGAGAGCGGCAATGCCCTTACCCAGAGGAAACCTGCCAGTGAAC 3474  
Db 3629 CAAGGAGAGAGCGGCAATGCCCTTACCCAGAGGAAACCTGCCAGTGAAC 3688





QY 1444 GATGGAAACAGACAGAGTCAAGAACTATCTTGGCACTGGCTGGGGAAGGTGGAAAGCCA 1503  
DB 2028 GATGGAAACAAACCAAGGTCAAGTCTATCTCTCGTTTGGCTGGGAAAGGTGGAAAGACC 2087  
QY 1504 TGTCCCTTAGTCAGGCTCTCCAAAGAGCATCTTTGTGTAAATGACATCTCTGTATGACAG 1563  
DB 2088 TGCCCTAGTAGTCAGGAATCTCCAGAGTACCGCTATGCATGATCATCTCTGTACACAG 2147  
QY 1564 CTTTACTGGAGACATCGCTTGGGCTTGGGCTTGTGTGTAGGACACATGCTTAACTGCCCTT 1623  
DB 2148 CTCTACTGGAGACATCTCTTGGGCTTGGGCTTGTGTGTAGGACACATGCTTAACTGCCCTT 2207  
QY 1624 AATGCAACATTTGGCTTGGAAATGGAAGCCAGTGTGTGTAGGACATTCAGACTCGGAGA 1683  
DB 2208 AATGTGACCATTTGGCTTGGAAATGGAAGCCAGTGTGTGTAGGACATTCAGACTCGGAGA 2267  
QY 1684 GTCTTCTGTGTCAAGAGTCACTGGGACCAAGTAAATGACCAAAAGATGTCCAGATTTCTACT 1743  
DB 2268 GTCTTCTGTGTCAAGAGTCACTGGGACCAAGTAAATGACCAAAAGATGTCCAGACTTACA 2327  
QY 1744 CGACCTGAACCTGTGGCCCTCTTTTCTCCCATGCAAAAGAGTGTATGTGACTGCT 1803  
DB 2328 CGCCAGAACTGTGAACCTTTTCTCTCCCATGCAAAAGAGTGTATGTGACTGCT 2387  
QY 1804 TTCAGTGTAGTGAACCTTCCCAAGAGTGTGCCAAGCAGGAAATGCCACAGTAAACAG 1863  
DB 2388 TTCAGTGTAGTGAACCTTCCCAAGAGTGTGCCAAGCAGGAAATGCCACAGTAAACAG 2447  
QY 1864 TCTCGATACAGATCATCATCAAGAGCAGGCAATGAGGCCAGGAATGCCAGATACC 1923  
DB 2448 TCTCGATACAGATCATCATCAAGAGCAGGCAATGAGGCCAGGATGCCAGATACC 2507  
QY 1924 TTATATGAGGAGAGAGTGTGAAGATTTCTTGTGTCTGTATATCGTGGAGGCA 1983  
DB 2508 TTGTTTGAAGAAAGAGTGTGAAGATTTCTTGTGTCTGTATATCGTGGAGCG 2567  
QY 1984 CAGAAATGGAGCCCTTGCATTTAGTGCAGAGTCTGTCTGGCAGGGAATAACGGGCAGC 2043  
DB 2568 CAGAAATGGAGCTTGTATCTCTGTCTCAGAGTCTATCAGCAGGAGGAGAACAGGCACC 2627  
QY 2044 AGTGAAGCTGTGGAAGGGTGTACAAACAGAGCTGTCTATGATCTCTGTATGACAAAC 2103  
DB 2628 AGTGAAGCTGTGGAAGGGTGTACAAACAGAGCTGTCTATGATCTCTGTATGATTAAC 2687  
QY 2104 CGGTCAAGAGAAATGATGGAATTCCTCAAGCAGACAAACGGCATCGCTCTCTTGTGCAA 2163  
DB 2688 CAGTCTGCAGAAATGACAGAGTGTCTCAAGCAGATGATGCAATGCAATGCAATGCAAT 2747  
QY 2164 GAATGCAGTCTCCATGTGCAAGAGTGTGCACTTCACTGTGTGTCTCAAGTTTACGCC 2223  
DB 2748 GAATGCAGTCTCCATGTGCAAGAGTGTGCACTTCACTGTGTGTCTCAAGTTTACGCC 2807  
QY 2224 TGCTCCACGAACTGTGAAGCCCAAAAGTATGGCGGCGACAGCTCACAGGGAAGGACAGA 2283  
DB 2808 TGCTCCACGAACTGTGAAGCACTCAGATCAGAGGCGGCGAGCTTACAGGGAAGGACAG 2867  
QY 2284 AAGAGGGAATGCGAGATTTCTGACCTTTTACCTCTTAGTGGAGACAGAACTATGCTCT 2343  
DB 2868 AAGAGGGAATGCGAGATTTCTGACCTTTTACCTCTTAGTGGAGACAGAACTATGCTCT 2397  
QY 2344 TGTGTGAATTTATATCCCACTTATGAAACTGTGTGAACTGTGTGAACTGTGTGAACTGT 2403  
DB 2928 TGTGTACGTTTATGTCTCATCTTATGAAACTGTGTGAACTGTGTGAACTGTGTGAA 2987  
QY 2404 AAGAGGAGCTTACC --- GAGGACTGGGGTACAAAGCAGACAGCAAGAAATGTGGAGAA 2460  
DB 2988 AAGAGGAGCTTACC --- GAGGACTGGGGTACAAAGCAGACAGCAAGAAATGTGGAGAA 3047  
QY 2461 GGCCTCGCTTTGAGAGTAGCTTTCTGTGATAAAATGGAAGCTTTGACCCCTCC 2520  
DB 3048 GGCCTCGCTTTGAGAGTAGCTTTCTGTGATAAAATGGAAGCTTTGACCCCTCC 3107  
QY 2521 TTCTGACGAGCTCTGTGTTTACATTCAGAAAAATGTGTCTTCCCTGCCCATTTGATTGC 2580

DB 3108 TTCTGCAACAGCTCTGGTTTACATTCAGGAAGCATGTGTATCCCTGTCCGTTGATTGC 3167  
QY 2581 AAGTTAAGCGATTGGTCTAGTTGGGGTCTTGCAGTTCACTTGTGGAATTTGAGTGAGA 2640  
DB 3168 AAGCTAAGTGACTGGTCCAGCTGGGGTCTTGCAGCTCACTTGTGGGATTTGAGTGAGA 3227  
QY 2641 ATTTCGATCCAAATGGCTTAAAGAAAACTTCAATGGAGGACGACCATGTGCCAACTG 2700  
DB 3228 ATTTCGATCCAAATGGCTTAAAGAAAACTTCAATGGAGGACGACCATGTGCCAACTG 3287  
QY 2701 GATCTCAAGAAATCAGGTACATGAGGAGGAGTCCCATGTTTACAGTGAGTCAATCAGTATTC 2760  
DB 3288 GATCTCAAGAAATCAGGTACATGAGGAGGAGTCCCATGTTTACAGTGAGTCAATCAGTATTC 3347  
QY 2761 TGGGTTGTAGAACACTGGTCTTCAATCAAAATCAACAATGAGCTGAGGTCCTTCGCTGT 2820  
DB 3348 TGGGTTGTAGAACACTGGTCTTCAATCAAAATCAACAATGAGCTGAGATCCCACTGT 3407  
QY 2821 GGAGGAGGACAACTTATAGGAATCAGATGTGTGAATCTTCCGATGCTGAGAGTGA 2880  
DB 3408 GGAGGAGGACAACTTATAGGAATCAGATGTGTGAATCTTCCGATGCTGAGAGTGA 3467  
QY 2881 GCAGTGTAGTGAACCTTGTGCAACAGGATGAAATTCCTCCAGAAACCCAGCTCTTCT 2940  
DB 3468 GCAGTGTAGTGAACCTTGTGCAACAGGATGATGTCTCCCAAGAAACCCAGGCTTCT 3527  
QY 2941 CTTATGTCTCCAAATGAGTGTGTCTGATGTGGGACTTTGGAGCAAAATGCCACAG 3000  
DB 3528 CTTATGTCTCCAAATGAGTGTGTCTGATGTGGGACTTTGGAGCAAAATGCCACAG 3587  
QY 3001 TCATGCGATCCCAACAATGCAAGAGAACTTCGCCACTCTCTAAGACCATCAGTGAAC 3060  
DB 3588 TCATGCGATCCCAACAATGCAAGAGAACTTCGCCACTCTCTAAGACCATCAGTGAAC 3647  
QY 3061 TCAGGACTTGTCTCAAGACTCAGAGTGCAGCTTGTCTCTGAAATGAAATTTCTTTC 3120  
DB 3648 TCAGGACTTGTCTCAAGACTCAGAGTGCAGCTTGTCTCTGAAATGAAATTTCTTTC 3707  
QY 3121 CAGTTCAGTGTACATCTAACAGAGTGGAGACATGCCAGCTGAGTGAAACCCAGCTCT 3180  
DB 3708 CAGTTCAGTGTACATCTAACAGAGTGGAGACATGCCAGCTGAGTGAAATTTCTTTC 3767  
QY 3181 GGTCAAGGCTCAGGACCCGCTTGAAGCTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 3240  
DB 3768 GGTCAAGGCTCAGGACCCGCTTGAAGCTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 3827  
QY 3241 ATGGAACATGTGAGCAGCATTAATTTGGAGAGCCCAAGAGATGAGCATTTCCCTGCTTG 3300  
DB 3828 ATGGAACATGTGAGCAGCATTAATTTGGAGAGCCCAAGAGATGAGCATTTCCCTGCTTG 3887  
QY 3301 GTGGAATGGTGTGCTCAACTGTGAGCTTCCAGGCTGAGCGCTTGGACAGAGTGTTCACAG 3360  
DB 3888 GTGGAATGGTGTGCTCAACTGTGAGCTTCCAGGCTGAGCGCTTGGACAGAGTGTTCACAG 3947  
QY 3361 ACCTGTGGCATGGAGGTGGAATGAGCCGGAATCTGATTTATCATTTATGCAACCCAGGA 3420  
DB 3948 ACCTGTGGCATGGAGGTGGAATGAGCCGGAATCTGATTTATCATTTATGCAACCCAGGA 4007  
QY 3421 GAAGGACGCGCATGCCCCCAGAGCTTACCCAGGAGAAACCTGCCAGTGAACCCCTGC 3480  
DB 4008 GAAGGACGCGCATGCCCCCAGAGCTTACCCAGGAGAAACCTGCCAGTGAACCCCTGC 4067  
QY 3481 TACAGTGGCTCTTGGCAACTGGTGTGATGTAAATTTGGAGGGTGGAGACTGTGGGAA 3540  
DB 4068 TATAGTTGGCTCTTGGCAACTGGTGTGATGTAAATTTGGAGGGTGGAGACTGTGGGAA 4127  
QY 3541 GAGTTTCAGATCCGAGCCCTTTCTGATGTGCTGAGTGTGCTCAATATCTCATGAGCT 3600  
DB 4128 GAGTTTCAGATCCGAGCCCTTTCTGATGTGCTGAGTGTGCTCAATATCTCATGAGCT 4187  
QY 3601 GAGTTTCAGATCCGAGCCCTTTCTGATGTGCTGAGTGTGCTCAATATCTCATGAGCT 3660



```
Db 4188 GTCCCTCTGGAGGAGCCCTGTGTGGAGAGTAGTACCCTTCCAGAGAGAAATCCCTGAAGCAG 4247
2Y 3661 CTGTGTTCTTGCTGCTCCAGGAGACTGTCATTTAAACAGATGGTTCAGAGTGGAGCACA 3720
2b 4248 CTGTGCTCTGTGCTTGCTCCAGGAGACTGCCACATACCCCATGGTCTGAGTGGAGCAG 4307
2Y 3721 TGTGAATTAACCTGTCATGTAGAGAGAGCTTTGAGACTGTGGCGGCCAGTCTAGATCA 3780
Db 4308 TGTGAGTAACCTGTCATGTAGAGAGAGCTTTGAGACTGTGGCGGCCAGTCTAGATCA 4367
2Y 3781 AGGACTTTTATAATTCAGTCTTTTGAAGACCAAGACAGCTGCCCCCAACAGAGTCTTAGAA 3840
Db 4368 AGGACTTTTATAATTCAGTCTTTTGAAGACCAAGAGTAGCTGCCCCCAACAGAGTCTTAGAA 4427
2Y 3841 ACAGCCCTTTGTACAGAGGAGCAATGTATTCATACATGAGAGAGCAAGTCTTTGGAAC 3900
2b 4428 ACAGGACCATGTACAGGTGGCAATGTATTCATACATGAGAGAGCAAGTCTTTGGAAC 4487
2Y 3901 AATAACGAACGAACGTATGTGTCAGCTGTCAGTGGCTTACATGGCGTTAATGTACAGAGGCTGC 3960
Db 4488 AATAATGAGCGGACTGTGTGTCAGCTGTCAGTGGCGTTAATGTACAGAGGCTGC 4547
2Y 3961 TCCTCTCAGCCCGTCTGCTGCTGCTATTCGGCAATGCAATCCAGCATGCAAAAACCTTTC 4020
Db 4548 TCCTCTCAGCCCGTCTGCTGCTGCTATTCGGCAATGCAATCCAGCATGCAAAAACCTTTC 4607
2Y 4021 TCCTACTGTACAGAGGTGAGTGTGTGTGAGAGGGCTATACAGAGATTAAGAA 4080
Db 4608 TCCTACTGTACAGAGGAGGAGTGTGTGTGAGAGGGATACACAGAGATTAAGAA 4667
2Y 4081 TCAATAGTTTCTCGTATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4140
Db 4668 TCAATAGTTTCTCGTATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4727
2Y 4141 GTGAAACCTTTCTGGAAGAAACAGACCTGTGAATTCAGAGATGATGATGATGATGATGATGAT 4200
Db 4728 GTGAAACCTTTCTGGAAGAAACAGACCTGTGAATTCAGAGATGATGATGATGATGATGATGAT 4787
2Y 4201 GGATGTTCTTCAACACCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4260
Db 4788 GGATGTTCTTCAACACCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4847
2Y 4261 TCAGTGGCGCTTTCTCATCATGATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 4320
Db 4848 TCAGTGGCGCTTTCTCATCATGATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 4907
2Y 4321 AAGCCAAACCATCATCAAGACACCTCCCAAGAGAGCTGATGATGATGATGATGATGATGATGAT 4380
Db 4908 AAGCCAAACCATCATCAAGACACCTCCCAAGAGAGCTGATGATGATGATGATGATGATGATGAT 4967
2Y 4381 GGAGACTTAGACATGTAA 4398
2b 4968 GGAGACTTAGACATGTAA 4985
```

```
RESULT 3
BC033125
LOCUS 2421 bp mRNA linear PRI 27-JUN-2002
DEFINITION Homo sapiens, clone IMAGE:3028838, mRNA, partial cds.
ACCESSION BC033125
VERSION BC033125.1 GI:21620030
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2421)
Strasbourg, R.
Direct Submission
Submitted (25-JUN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
```

REMARK  
COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgabbs-re@mail.nih.gov](mailto:cgabbs-re@mail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Genome Sequence Centre,  
BC Cancer Agency, Vancouver, BC, Canada  
[info@bcgsc.bc.ca](mailto:info@bcgsc.bc.ca)  
Steven Jones, Jennifer Asano, Ian Bosdet, Varon Butterfield,  
Suaanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,  
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo  
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven  
Nees, Pawan Pandoh, Anna-Lilisa Prabhu, Parvaneh Saeedi, Jacqueline  
Schein, Duane Smalhus, Michael Smith, Lorraine Spence, Jeff Stott,  
Michael Thorne, Miranda Tsai, Natasja van den Bosch, Jill Vardy,  
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IBAL Plate: 43 Row: a Column: 6  
This clone was selected for full length sequencing because it  
passed the following selection criteria: Hexamer frequency ORF  
analysis.

FEATURES  
source

Location/Qualifiers  
1..2421  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3028838"  
/tissue type="Muscle; rhabdomyosarcoma"  
/clone\_id="NTH MGC 17"  
/lab\_host="DH10B-R"  
/note="Vector: pOTB7"  
<1..2224  
/codon\_start=2  
/product="Unknown (protein for IMAGE:3028838)"  
/protein\_id="AAH33125.1"  
/db\_xref="GI:21620031"

## CDS

/translation="PHLKEINPSORTVLDPNSDSNERVTFKHOSYKAHHHSKSHALBI  
GYTRQVSTKSDGNAMSLCLQDSPLTVQSCIMPDCDTSOWSNWSKSCSKTRSG  
SLLPFRSRNRNVKHAIGGKECPLEKAEQIVGELLQCCPRYSWRVSEMKCVQ  
SLLEQDPHWHVTGPGVGGGQITREYCAQSVPAALAAKAEVREKALCVGPAP  
LPOLCNIPTGSDICVSWAGWLCIHENCHDPOGKGFRTORHVLMESTGPAGHCP  
HLVSPVEDPCVYRWLASGICPPDHGKGLGHLRIKAVCONDGRGDSVSLCPVPP  
PPEKSKCEIPEMDVILSEMTSSSCSQSNKNSDKQTRSTLILALAGEGKPCPP  
SQALQERHLNDHSCMLHWEISFWGPCSEDILVLTALNATIGNGENATCGVGTQKRV  
FCVSHRGQVMTKRCPTETVPRCPFLPKDCIVTAFSEWTPCPMCOAGNATVK  
QSVYRIITQEAANGQECPTLYEERECEDVSLCPVYRWKPKQWSPCLIPESVWQVI  
TGSSEACGKGLQTRAVSCI8DNRSAEMECLEKQTNMGMLLVQETVPCREDCTFTAW  
SKFTPCSTNCEATKSRRLTGKSRKEKQCDGLPLVETELCPDCEFTSQPVGNWS  
DCILPEGRREPHRLGLRVQADSKEGCEGLRFRAVACSKNGRPPVDPSPSCSSSGKEMDGE  
"

## ORIGIN

Query Match 50.0%; Score 2200.2; DB 9; Length 2421;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2202; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 336 GCCTCATCTTAAAGAAATTAATCCAGCGGAGAACTGTTCTTGATTTTAACCTGATTC 395  
Db 1 GCCTCATCTTAAAGAAATTAATCCAGCGGAGAACTGTTCTTGATTTTAACCTGATTC 60  
QY 396 AAATGAGCGAGTACCTTTAAACATCAAGTTACAAAGCACATCATCATTCGAAAGTCTTG 455  
Db 61 AAATGAGCGAGTACCTTTAAACATCAAGTTACAAAGCACATCATCATTCGAAAGTCTTG 120  
QY 456 GGCAATGAGATAGTGTATCAACCCGCGAGGTTTCGTGTACAAAGATGATGACAAAA 515  
Db 121 GGCAATGAGATAGTGTATCAACCCGCGAGGTTTCGTGTACAAAGATGATGACAAAA 180  
QY 516 TGCTATGTAAAGCTTTGCCTTTCAAGATTCTCCCATTTGACTGTTCCAGTCTCTCATCAT 575

Db 181 TGCTATGTTAAGCCTTTGCCCTTCAAGATTTCCTCCATTGACTGTTTCAGTCCCTCATCAT 240  
QY 576 GCCCAAGACTGTGAACCTCCAGTGTCTCTCTGGAGCCCTGCTCCCAAGACATGCCG 635  
Db 241 GCCCAAGACTGTGAACCTCCAGTGTCTCTCTGGAGCCCTGCTCCCAAGACATGCCG 300  
QY 636 TTCAGGGAGTCTCTGCGAGGATTTAGGAGCAGGAGCCGGAACGTGAAGCAGATGGCTAT 695  
Db 301 TTCAGGGAGTCTCTGCGAGGATTTAGGAGCAGGAGCCGGAACGTGAAGCAGATGGCTAT 360  
QY 696 TGGAGGTGAAGAGGAGTGTCTGTAACCTTTTCAGAAAAGAGCCCTGCATTGTTGAAGGAGA 755  
Db 361 TGGAGGTGAAGAGGAGTGTCTGTAACCTTTTCAGAAAAGAGCCCTGCATTGTTGAAGGAGA 420  
QY 756 ACTTCTGAGCAATGTCCTCCAGTATTCCTGAGAACTTCTGAATGGAAGAAATGCCAAGT 815  
Db 421 ACTTCTGAGCAATGTCCTCCAGTATTCCTGAGAACTTCTGAATGGAAGAAATGCCAAGT 480  
QY 816 CTCTCTCTCTCGAGCAGCAGGATCCCACTGCGCATGTGACGGGACCCGTTGTGGCGG 875  
Db 481 CTCTCTCTCTCGAGCAGCAGGATCCCACTGCGCATGTGACGGGACCCGTTGTGGCGG 540  
QY 876 TGGGATCCAGACCCGGGAGGTGTAATGTGCCAGAGCGTACAGCAGCTGCCGCACTGAG 935  
Db 541 TGGGATCCAGACCCGGGAGGTGTAATGTGCCAGAGCGTACAGCAGCTGCCGCACTGAG 600  
QY 936 GGCCAAGGAAGTCTCTAGACCTGTGGAAGGCAATTATGTGGGACCCGCCGTTGTC 995  
Db 601 GGCCAAGGAAGTCTCTAGACCTGTGGAAGGCAATTATGTGGGACCCGCCGTTGTC 660  
QY 996 CTCTCAGCTCTGCAATATCCCTTGCTCTACGAGTGCATAGTATCTTCTGCTGAGCCTG 1055  
Db 661 CTCTCAGCTCTGCAATATCCCTTGCTCTACGAGTGCATAGTATCTTCTGCTGAGCCTG 720  
QY 1056 GGGCCCTGTGCATCCATGAATCTGATGAACCTCAGGGGAAAAGATTTAGAACGAG 1115  
Db 721 GGGCCCTGTGCATCCATGAATCTGATGAACCTCAGGGGAAAAGATTTAGAACGAG 780  
QY 1116 GCAGCGCATGTCTCTATGGAATCTACAGGCGCTGCAGGCGATTGCCCTCATTTTGGTGA 1175  
Db 781 GCAGCGCATGTCTCTATGGAATCTACAGGCGCTGCAGGCGATTGCCCTCATTTTGGTGA 840  
QY 1176 GTCTGTTCTGTGAGGATCCATGTGCTACGATGGCTGCATCAGAGGATCTGTTT 1235  
Db 841 GTCTGTTCTGTGAGGATCCATGTGCTACGATGGCTGCATCAGAGGATCTGTTT 900  
QY 1236 CCTCATCATGAAAATGTGGCTGGGACATCGTATTCTGAAGCGCTGCTGCCAGATGA 1295  
Db 901 CCTCATCATGAAAATGTGGCTGGGACATCGTATTCTGAAGCGCTGCTGCCAGATGA 960  
QY 1296 CCGCGGAGAGATGTATCAGGAGTCTTTGCCAGTTCCTCCCTGCTCTGAGAGGAATC 1355  
Db 961 CCGCGGAGAGATGTATCAGGAGTCTTTGCCAGTTCCTCCCTGCTCTGAGAGGAATC 1020  
QY 1356 TTGTGAATTCCTCCGAATGCACTGTGTGTGAGCGAGTGGACGGAGTGTCTATCCTG 1415  
Db 1021 TTGTGAATTCCTCCGAATGCACTGTGTGTGAGCGAGTGGACGGAGTGTCTATCCTG 1080  
QY 1416 TTCCAGTCTCTTTCAATAAAAACTAGATGGGAAAACAGACCGGTCAAGAACTATCCT 1475  
Db 1081 TTCCAGTCTCTTTCAATAAAAACTAGATGGGAAAACAGACCGGTCAAGAACTATCCT 1140  
QY 1476 GGCATGCTGGGAGGTGGAAAGCAATGTCCTCCCTAGTCAGGCTCTCCAAAGCATCG 1535  
Db 1141 GGCATGCTGGGAGGTGGAAAGCAATGTCCTCCCTAGTCAGGCTCTCCAAAGCATCG 1200  
QY 1536 TTGTGTAAATGACCATTCCTGTATCAGCTTCACCTGGGAGACATCGCCTTGGGCGCCCTTG 1595  
Db 1201 TTGTGTAAATGACCATTCCTGTATCAGCTTCACCTGGGAGACATCGCCTTGGGCGCCCTTG 1260  
QY 1596 TTCTGAGGACATTTGGTAACTGGCCCTTAATGCAACCATTTGGCTGGATGAGAGGCAAC 1655  
Db 1261 TTCTGAGGACATTTGGTAACTGGCCCTTAATGCAACCATTTGGCTGGATGAGAGGCAAC 1320

QY 1656 GTGTGTGTAGGCAATTCAGACTCGGAGAGTCTTCTGTGTCAAGAGTCAAGTGGGCAAGT 1715  
Db 1321 GTGTGTGTAGGCAATTCAGACTCGGAGAGTCTTCTGTGTCAAGAGTCAAGTGGGCAAGT 1380  
QY 1716 AATGACCAAAAGATGTCAGATTCTACTCGACTGAAACTGTGCGCCCTGTTTCTCCC 1775  
Db 1381 AATGACCAAAAGATGTCAGATTCTACTCGACTGAAACTGTGCGCCCTGTTTCTCCC 1440  
QY 1776 ATGCAAAAAGAGTGTATTTGTGACTGCTTTTCAGTGTAGTGGACACCTGCCCCAAGGATGTG 1835  
Db 1441 ATGCAAAAAGAGTGTATTTGTGACTGCTTTTCAGTGTAGTGGACACCTGCCCCAAGGATGTG 1500  
QY 1836 CCAAGCAGAAATGCCACAGTAAACAGTCTTCGATACAGAAATCATCATCAAGAGAGCAGC 1895  
Db 1501 CCAAGCAGAAATGCCACAGTAAACAGTCTTCGATACAGAAATCATCATCAAGAGAGCAGC 1560  
QY 1896 CAATGAGGCCAGGAATGCCAGATACCTTATATGAGGAGAGAGAGTGTGAAGATGTTTC 1955  
Db 1561 CAATGAGGCCAGGAATGCCAGATACCTTATATGAGGAGAGAGAGTGTGAAGATGTTTC 1620  
QY 1956 CTTGTGCTCTGTATATCGGTGGAAGCCACAGAAATGAGGCCCTTGCACTCTTAGTCCAGA 2015  
Db 1621 CTTGTGCTCTGTATATCGGTGGAAGCCACAGAAATGAGGCCCTTGCACTCTTAGTCCAGA 1680  
QY 2016 GTCTGTCTGCGAGGGAATAAACGGGAGAGTGAAGCTGTGGAAGGGGTTACAACCAAG 2075  
Db 1681 GTCTGTCTGCGAGGGAATAAACGGGAGAGTGAAGCTGTGGAAGGGGTTACAACCAAG 1740  
QY 2076 AGCTGTCTCATGCACTCTCTGTATGACAAACCGGTGACAGAAATGATGGAATGCTCAAGCA 2135  
Db 1741 AGCTGTCTCATGCACTCTCTGTATGACAAACCGGTGACAGAAATGATGGAATGCTCAAGCA 1800  
QY 2136 GACAAACGGCATGCTCTCTGTGTCNAGAAATGACAGTCCCATCTCGAGAGAGCTGCAC 2195  
Db 1801 GACAAACGGCATGCTCTCTGTGTCNAGAAATGACAGTCCCATCTCGAGAGAGCTGCAC 1860  
QY 2196 CTTCACTGCTTGGTCCAGTCTTACGCCCTGCTCCAGAACTGTGAAGCCACAAAAGTAG 2255  
Db 1861 CTTCACTGCTTGGTCCAGTCTTACGCCCTGCTCCAGAACTGTGAAGCCACAAAAGTAG 1920  
QY 2256 GCGGCGACAGCTCACAGGAAAAGCAGAAAGAGAGAAATGCCAGGATCTGACCTTTA 2315  
Db 1921 GCGGCGACAGCTCACAGGAAAAGCAGAAAGAGAGAAATGCCAGGATCTGACCTTTA 1980  
QY 2316 CCTCTAGTGGAGACAGAACTATGCTTGTGATGAAATTTATATCCCACTTATGGA 2375  
Db 1981 CCTCTAGTGGAGACAGAACTATGCTTGTGATGAAATTTATATCCCACTTATGGA 2040  
QY 2376 CTGCTCAGATTGCTTCTCCAGAGGAGGAGGCTCACCAGGAGCTGCGGGTACA 2435  
Db 2041 CTGCTCAGATTGCTTCTCCAGAGGAGGAGGCTCACCAGGAGCTGCGGGTACA 2100  
QY 2436 AGCAGACAGCAAGAAATGTGGAGAGGCTGCGCTTTCGAGCAGTAGCTGTTCTGATAA 2495  
Db 2101 AGCAGACAGCAAGAAATGTGGAGAGGCTGCGCTTTCGAGCAGTAGCTGTTCTGATAA 2160  
QY 2496 AAATGGAAGACCTGTTGACCCCTCTTCTGACAGAGCTCTGTTA 2540  
Db 2161 AAATGGAAGACCTGTTGACCCCTCTTCTGACAGAGCTCTGTTA 2205

## RESULT 4

BC019344  
LOCUS  
DEFINITION  
Homo sapiens KIAA1679 protein, mRNA linear PRI 04-NOV-2003  
partial cds.  
ACCESSION  
BC019344  
VERSION  
BC019344.1 GI:17939567  
KEYWORDS  
Homo sapiens (human)  
Homo sapiens  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;



Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 1 (bases 1 to 2422)  
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shermen, C.M., Schuler, G.D., Altschul, S.F., Zebberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, P., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stappleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodrigues, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalls, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.  
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
 22388257  
 12477932  
 2 (bases 1 to 2422)  
 Strausberg, R.L.  
 Direct Submission  
 Submitted (13-DEC-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada  
 info@bcsc.bc.ca  
 Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Letticia Heiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Liisa Prabh, Parvaneh Saeedi, Jacqueline Schein, Duane Smalls, Michael Smith, Lorraine Spence, Jeff Scott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAL Plate: 18 Row: e Column: 12.  
 Features Location/Qualifiers  
 1. .2422  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3530946"  
 /tissue\_type="Muscle, rhabdomyosarcoma"  
 /clone\_lib="NIH\_MGC\_17"  
 /lab\_host="DH10B-R"  
 /note="Vector: pOTB7"  
 <1. .2422  
 /gene="KIAA1679"  
 /db\_xref="LocusID:80731"  
 <1. .2224  
 /gene="KIAA1679"  
 /codon\_start=2  
 /product="KIAA1679 protein"  
 /protein\_id="AAH19344.1"  
 /db\_xref="GI:17939568"  
 /db\_xref="LocusID:80731"

/translation="PHLKEINPSGRVTLDPNSDSNERVTFKHOSYKAHHHSKWAIEI  
 GYTRQVSCSTRSDGNAMLSLCLQDQFPPLTVQSCIMPCKCETQSWSPSCSTCSG  
 SLPLPGFASRNKVMKMAIGGKECEACIVGELIQCCPRTYVSWRTSEWKEQV  
 LLLEODPHHWVTGPGVGGIOTREYCAQSVPAALAAKAEVSRPVEKALCVGAP  
 LLSOLNCPGCTDCIVSSWSAGLCHENCHDPQKKGPRTRORHVLVSTGAGHCP  
 HLVSVPCEPDMCYRLASGICFPDQKGGKGLGHRLLKAVCONDREDVSGSLCPVPP  
 PPKAKSCEIFCRMDCVLSEWNSQSCSNKNSDGKQTRSTRILLALAGEGKFPFP  
 SQAQHEILCNHDSQMLHWETSPWGPSCSTNLTALNATIGNWGATCGVGIQTRRV  
 FVKVSHQVMTKRCPCDSTPRTFCFKKDCIVTAFSEWTPCPRMCOAGNATVK  
 QSYRIIOEAANGQCEPDLYEERECEDVSLCPYVRMKPKWSPCILLVPEVWQI  
 TGSEACGKGLQTRAVSCISIDNRSKAEMLKQTMGPLLVOEICTVPCREDCTFTAW  
 SKFTPCSTNCEATKSRRLQTKSRKKEQSDLYPLVETELCPDEFTSQPYGNWS  
 DCILPEGRPREHRLRVQADSKEGELRPRVACSDKNRPRVDPSPFSCSSSGKEMDE  
 "

ORIGIN

Query Match 50.0%; Score 2200.2; DB 9; Length 2422;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 2202; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 336 GCCTCATCTTAAGAAATTAATCAACGCGAAGAACTGTTCTCGATTTTAACTCTGATTC 395  
 DB 1 GCCTCATCTTAAGAAATTAATCAACGCGAAGAACTGTTCTCGATTTTAACTCTGATTC 60  
 QY 396 AAATGAGCGAGTCACCTTTAAACATCAAGTTACAAAGCACATCATCATTCGAGTCTTG 455  
 DB 61 AAATGAGCGAGTCACCTTTAAACATCAAGTTACAAAGCACATCATTCGAGTCTTG 120  
 QY 456 GCGAATAGAGATAGTTATCAACCCGCGAGTTCTGTGACAGAAAGTATGACAAAG 515  
 DB 121 GCGAATAGAGATAGTTATCAACCCGCGAGTTCTGTGACAGAAAGTATGACAAAG 180  
 QY 516 TGCTATGTTAAGCCTTTGCGCTTCAAGATTCCTCCCATTCGACTGTTCCAGTCTGCATCAT 575  
 DB 181 TGCTATGTTAAGCCTTTGCGCTTCAAGATTCCTCCCATTCGACTGTTCCAGTCTGCATCAT 240  
 QY 576 GCCCAAGAGCTGTGAACCTCCAGTGGTCTCTCTCGAGCCCTCTGCTCCAAGACATGCCG 635  
 DB 241 GCCCAAGAGCTGTGAACCTCCAGTGGTCTCTCTCGAGCCCTCTGCTCCAAGACATGCCG 300  
 QY 636 TTCAGGAGTCTCTTGGCAGGATTTAGGACGAGCGGGAAGTGAAGCACATGGCTAT 695  
 DB 301 TTCAGGAGTCTCTTGGCAGGATTTAGGACGAGCGGGAAGTGAAGCACATGGCTAT 360  
 QY 696 TGGAGGTGGAAAGAGTGTCTCTGAACCTTTTGAAGAAAGAGCCCTGCTGTTGAAGAGAGA 755  
 DB 361 TGGAGGTGGAAAGAGTGTCTCTGAACCTTTTGAAGAAAGAGCCCTGCTGTTGAAGAGAGA 420  
 QY 756 ACTTCTGAGCAATGTCCAGGATTCCTGAGAACTTCTGAAATGGAAGAAATGCCAAGT 815  
 DB 421 ACTTCTGAGCAATGTCCAGGATTCCTGAGAACTTCTGAAATGGAAGAAATGCCAAGT 480  
 QY 816 CTCTCTCTCTCGAGCAGCAGGATCCCACTGTCATGTGACGGACCCCTGTGTGGCGG 875  
 DB 481 CTCTCTCTCTCGAGCAGCAGGATCCCACTGTCATGTGACGGACCCCTGTGTGGCGG 540  
 QY 876 TGGATCCAGACCCGGAGGTGTACTGTGCCAGAGCGTACCAAGCAGCTGCCGACATGAG 935  
 DB 541 TGGATCCAGACCCGGAGGTGTACTGTGCCAGAGCGTACCAAGCAGCTGCCGACATGAG 600  
 QY 936 GGGCAAGGAGTCTTAGACCTGTGGAAGAGGCAATTATGTGGGACCCGCCCGTTGCC 995  
 DB 601 GGGCAAGGAGTCTTAGACCTGTGGAAGAGGCAATTATGTGGGACCCGCCCGTTGCC 660  
 QY 996 CTCTCAGCTCTGCAATATCCCTTCTCTAGGACTGCATAGTATCTTCTCTGGTCTGAGCTG 1055  
 DB 661 CTCTCAGCTCTGCAATATCCCTTCTCTAGGACTGCATAGTATCTTCTCTGGTCTGAGCTG 720  
 QY 1056 GGGCTGTGCAATCCATGAACCTGTGAACTGAGACCTCAGGGGAAAAAGAGATTAGAACGAG 1115  
 DB 721 GGGCTGTGCAATCCATGAACCTGTGAACTGAGACCTCAGGGGAAAAAGAGATTAGAACGAG 780  
 QY 1116 GCAGGCGCATGTCTCATGGAATCTACAGGCGCTGCAGGGCAATTCGCCCTCATTTGGTGGGA 1175

1861	CTTCACCTGCTTGGTCCAGTTTACGGCCCTGCTCCAGCAACTGTGAAGCCACAAAAGTAG	1982
2256	CGCGGCAGAGCTCACAGGGGAAAAAGCAGAAAGAGGAGAAATCCACAGGATTCGTACCTTTTA	2315
1921	CGCGGCAGAGCTCACAGGGGAAAAAGCAGAAAGAGGAGAAATGCCAGGATTCGTACCTTTTA	1980
2316	CCCTCTAGTGGAGACAGAACTATGCTTGTGATGTAATTTATATCCACACCTTATGGAAA	2375
1981	CCCTCTAGTGGAGACAGAACTATGCTTGTGATGTAATTTATATCCACACCTTATGGAAA	2040
2376	CTGTGTCAGATTGCAATTCCTCCAGAAGGCGAGAGGAGGAGCCCTCACCGAGGAGCTCGCGGTACA	2435
2041	CTGTGTCAGATTGCAATTCCTCCAGAAGGCGAGAGGAGGAGCCCTCACCGAGGAGCTCGCGGTACA	2100
2436	AGCAGACAGCAAGAAATGTGGAGAAGGCGCTGGCTTTCGAGCAGTAGCTGTTCTGTGATAA	2499
2101	AGCAGACAGCAAGAAATGTGGAGAAGGCGCTGGCTTTCGAGCAGTAGCTGTTCTGTGATAA	2160
2496	AAATGGGAAGACCTGTTGACCCCTCCCTTTCGACGAGCTCTGGTTA	2540
2161	AAATGGGAAGACCTGTTGACCCCTCCCTTTCGACGAGCTCTGGTTA	2205
<p>RESULT 5</p> <p>AB023177 Homo sapiens mRNA for KIAA0960 protein, partial cds.</p> <p>LOCUS AB023177.2 GI:20521709</p> <p>DEFINITION Homo sapiens (human)</p> <p>ACCESSION AB023177</p> <p>VERSION AB023177.2</p> <p>KEYWORDS</p> <p>SOURCE</p> <p>ORGANISM</p> <p>Homo sapiens</p> <p>Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.</p> <p>REFERENCE</p> <p>AUTHORS</p> <p>Nagase, T., Ishikawa, K., Suyama, M., Kikuno, R., Hirose, M., Miyajima, N., Tanaka, A., Kotani, H., Nomura, N. and Ohara, O.</p> <p>TITLE</p> <p>Prediction of the coding sequences of unidentified human genes. XIII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro</p> <p>JOURNAL</p> <p>MEDLINE</p> <p>99246063</p> <p>10231032</p> <p>2 (bases 1 to 5669)</p> <p>Chara, O., Nagase, T. and Kikuno, R.</p> <p>REFERENCE</p> <p>AUTHORS</p> <p>Submitted (04-FEB-1999) Osamu Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba 292-0812, Japan. (E-mail: cdnainfo@kazusa.or.jp, Tel:+81-438-52-3913, Fax:+81-438-52-3914)</p> <p>JOURNAL</p> <p>On May 9, 2002 this sequence version replaced gi:4589563.</p> <p>COMMENT</p> <p>FEATURES</p> <p>Location/Qualifiers</p> <p>1..5669</p> <p>/organism="Homo sapiens"</p> <p>/mol_type="mRNA"</p> <p>/db_xref="taxon:9606"</p> <p>/clone="Hj057951"</p> <p>/sex="male"</p> <p>/tissue_type="brain"</p> <p>/clone_lib="pBluescriptII SK plus"</p> <p>/dev_stage="adult"</p> <p>/note="This sequence was obtained by subcloning of the DNA fragments derived from two cDNA clones (1 - 1401 was derived from fh03753 and 1402 - 5669 was derived from hj05779)."</p> <p>gene</p> <p>1..5669</p> <p>/gene="KIAA0960"</p> <p>1..4510</p> <p>/gene="KIAA0960"</p> <p>/codon_start=2</p> <p>/product="KIAA0960 protein"</p> <p>/protein_id="BAA76804.2"</p> <p>/db_xref="GI:20521710"</p> <p>CDS</p>		

/translation="RBIACIQKDKOIPABDII:CEYFEPKPLLEQAACLIPOCDQCIIVE  
FSAWSECTSCGSLQHRVHVAPPOFGSGCPNLTEFOVCS9PCEABELRYSLHV  
GWSCTSCMPHSROVRAARRKNKREKDRSGKVDPEARELLIKKRNRRNQROENK  
YNDIOLGYCTREVMCLNKTKGARDLSFCQOEXLPMTFOSCVITKEQVSEWSEWSPCS  
KTHCDVSPAGTRVTRIRQPIGSEKCEPEFEKPCLSQSDGVVPCATYGNWRITE  
WTECRVDPILLSQDQRGHTALCGGIGITREVIYQANENLLSULSTHKNKEASKPM  
DKLKTGPTIPNTQULCHIPTCEPEVPSWAMPCTYENCNDQOQKGFKLKRITIN  
EPTGSGVTCNGCPHLLLEAIPCEPAPCDWKLADGCEPDNGKEGCPGTQVOEVVCIIN  
SDEEVDROLCDRAIPPIVACDAPCPDCLVSTWSSCSHTCSGKTTEGQKIRAR  
SILAYAGERGLRCNPSSALQEVRSNEHPCTVHMOTGPMWQCTEDTSVSEFNTTIT  
WNGEASCSVMOTRKVIVCVUNVGOVGPKKCPESLRPETVRPCLLPKCKDCIVTEPSD  
WNSPSSCKEGRSSIKQSRHVIQLPANGARDCTDPLYERKAGAEFOACSINWKT  
HKWRCOLVPSVQSPQAGQEGPGRARAITCKQDGGQAGHECLOVAGYAPFAL  
TQACQIPCDQDQOLTSWFSFSCNGPCGAVRTRKTLVGSKKCKKKNCHLYPLIET  
QYCPCKYNAQPVGNMNDGILPEKGVILLGMKVQDGIKECGQYRYQAMACHYDNGR  
LVETSRCSNGHYIEEACIIPCSPSDKLSEWNSWRSKSGSVKVRKSWLREKPYNG  
GRPCPLDHVNOAQVVEVVPDSDNOYLWTEPMSICKVTFVNNRENGEGVQKRV  
RCMONTADGSEHEVDYLDPEMPLGSRVCKLPCEPDCVISEMGPWQCVLPQNQSS  
FRORSADPQPADEGRSEPNAVEKEPCLNKNCHYDINVDWSTCQLSEKAVCGNG  
IKTRMLDCVRSKGSVDLKYCEALGUEKQWQNTSCMECPVNCQSDMSPSECSQT  
CELTGMIRRTVTPQDGRPCPSLMDQSKPCPVKCYRWQYQWSPQVQEAQCG  
EGTRTRNISCVDGADDFSKVVDDEFADIELIIDGNKMLVLESQSPQCPGDCYL  
KXMSWSLQOLTVCNVEDLFGGIIQVRSRPVIOELNOHLCPBOMLETKSKYDQCY  
EYKMASANKGSRSTVMKORSQDINVTGCLVMSQDADRNCNPPCSPHYSYCSRTKI  
CHCEEGYEVMSNSNLEACTLIPVVLPTMEDKGDVKTSAVHPHTQPSNPAGRG  
TWLFQFPGPGRUKTWYVGAAGFVLLIFIVSMYIACKPKPKPQRRQNNRKLPLTL  
AYDGDADM"

ORIGIN

Query Match 28.0%; Score 1232.8; DB 9; Length 5669;

Best Local Similarity 57.7%; Pred. No. 0;

Matches 2598; Conservative 0; Mismatches 1782; Indels 126; Gaps 17;

2y	10	TGCAATCAGAACTGACCGAACTGTGGTTGGCAATGAAATATGCAACACTTTGCCCTT	69
2b	14	TGCATCAGAAAGACAAAGACATCTCTCGCGAGGATATCATCTGTGAGTACTTTGAGCC	73
2y	70	CAGCTCTACAGACAGGCTGCTCATCTCTGTCCTCGGATGTGTAGTATCTGAG	129
2b	74	AAGCTCTCTGAGAGAGGCTGCTCATCTCTGTCCTCGGATGTGTAGTATCTGAG	133
2y	130	TTCTTACATGGTCCAACTGTAGCAAGGAGTGTGGGAAGAAATGACGATAGAACTCGC	189
2b	134	TTTTCTCTCTGCTCGAATGCTCAAGACCTGCGGACGGGCTCCAGCACCGGAGCGT	193
Qy	190	CGGTCATAGTCTCCCTCTCTTTGTGGTGTGGCAATGTGCAATCTGAGTCAAGA	249
Db	194	CATGTGGTGGCCCCCGGAGTTCGGAGGCTCTGGCTGTCCAACTGACGGAGTTCAG	253
Qy	250	GCCTGTGATGCTCCATTTCTGTCCTCTTGGGGAAGAGGATATACATTTAGCCTTAAG	309
Db	254	GTGTG-----CCAATCCAGTCCATGCGAGCGGCGAGGCTCAGGTACAGCTGCAT	304
Qy	310	GTGGACATGAGGATAATGAGACTGCTCATCTTAAAGAAATTAATCCAGCGGAAGA	369
Db	305	GTGGGGCCCTGGAGCACCTGCTCAATGCCCACTCCCGACAGTAAGACAAGCAAGGAGA	364
Qy	370	ACTGTTCTGGAATTTAA-----CTCTGATTCAAATGAG	402
Db	365	CGCGGAAGAAATAAGAACGGGAAGAGACCGGACGAGGATTAAGGATCAGAGGCC	424
Qy	403	CGAGTCACCTTTAAACATCAAGTTACA-----AAGCACATCATCTATTCGAAGTCT	453
Db	425	CGCGAGCTTTAAGAAAGAAAGAGAAACAGAAACAGGACAGAAACAGAGAGACAAATAT	484
Qy	454	TGGCAATAGAGATAGTTATCAACCGGAGGTTTCGTGTACAGAGATGATGACAA	513
Db	485	TGGGACATCCAGATTGGATATCAGACAGAGAGGTTATGTGCATTAACAAGA	544
Qy	514	AATGCTATGTTAAGCCTTTGCTTCAAGATTCCTTCCCATTTGACTGTTTCACTCTGCTC	573
Db	545	GCTGCTGATTTAGCTTTTGGCCAGCAGAGAGAGCTTCCATGACCTTCCAGTCTCTGTG	604

1685 AACACAACTACGACTTGAATGGGAGGCTCTCTCTGTGCGCATGCACACAGAAA 1744  
1684 GTCCTCTGTGTCAAGAGTCACTGTGGGACAAGTAATGACCAAAAGATGTCAGATTCTACT 1743  
1683 GTCATCTGTGTGCGAGTCAATGTGGGCAAGTGGGACCAAAATATGTCCTGAAGCCCTT 1804  
1744 CGACTGAACTGTGCGGCCCTGTCTTCCATGCAAAAAGACTGTATTGTCACTGCT 1803  
1805 CGACTGAACTGTGAAGCCCTGTCTGCTCTGTTGTAAGAGGACTGTATTGTGACCCCA 1864  
1804 TTCAGTGTAGTGACACCTGCGCCCAAGGATGTCGAAGCAGAGAAATGCCACAGT---AAAA 1860  
1865 TATAGTGTAGTGACATCATGCGCCCTCTCTGTTGTAAGAGGAGTCTCCAGTATCAGGAAG 1924  
1861 CAGTCTCGATACAGAAATCATCATCAAGAGCAGCAATGAGGCCAGGAATGCCAGAT 1920  
1925 CAGTCTAGGCATCGGGTCAATCATTCAGTGTGCGCAACACGGGGCCGAGACTGCACAGAT 1984  
1921 ACCTTATATAGGAGAGAGTGTGAAGATGTTCCCTGTGTCTGTATATATCGTGGAG 1980  
1985 CCCCTCTATGAAGAGAGCCCTGTGAGGACCTCAAGCGTGCCAAAGCTACAGTGGAG 2044  
1981 CCACAGAAATGAGGCCCTTGACCTCTAGTGTGCGCAGAGTCTGTCTGCGCAGGGAATAACGGGC 2040  
2045 ACTCACAATGGCGCAGATGCCAATTAGTCCCTTGGAGCGTGCAACAGCAGCCCTGGA 2104  
2041 AGCAGTGAAGCTGTGGAAGGGTTACAAACAGAGCTGTCTCATGCTCTGTATGATGAC 2100  
2105 GCACAGGAGGCTGTGGGCCCTGGGCGCAGCAGGCAAGAGCCATTACTTGTGCGCAAGCAAGAT 2164  
2101 AACCGGTGACAGAGAAATGATGGAATGCCCTCAAGCAGACAAACCGCATGCCCTCTCTTTGTG 2160  
2165 GGAGGACAGGCTGGAATCCATGAGTGCCCTACAGTATGCGAGGCCCTGTGCGCAGCCCTTACC 2224  
2161 CAAGAAATGCAAGTCCCATGTGCGAGAGACTGCACCTTCACTGTGTGGTCCAAATTAAG 2220  
2225 CAGGCTCCAGATCCCTGCGCAGGATGACTGTCAATTTGACCACTGGTCCAAATTTCT 2284  
2221 CCCTGCTCCAGAACTGTGAAGCCCAAAAGTAGTGGGCGCAGACAGCTCACAGGGAAGAGC 2280  
2285 TCATGCAATGAGACTGTGTGCGAGTATAGGACCAAGAGCCACTCTTTGTTGGAAGAAAGT 2344  
2281 AGAAAGAGGAGAAATGCCAGATTTGACCTTTACCTTTAGTGGAGACAGAACTATGT 2340  
2345 AAAAGAGAGGAAAAATGTAAAAATTTCCCATTTGTATCCCTGATGAGACTCAGTATTGT 2404  
2341 CTTGTGTATGATTTATATATCCCACTTATGGAAGTGTGCAAGTGTGATTTCCAGAA 2400  
2405 CTTGTGCAAAATATATGCAACCTGTGGGAGCTGGTCAAGACTGTATTTTTACCAGAG 2464  
2401 GGCAGAGGGAGCTTCAACCGAGGACTGCGGTTACAAGCAGACAGCAAGAAATGTGGAGAA 2460  
2465 GGAAGTGGAGTGTGCTGGAAATGAAGTACAAAGAGACATCAAGGAATGCGGACAA 2524  
2461 GGCCTGCGCTTTCGAGCAGTAGCTGTCTGATAAATGGAAGACCTGTGACCCCTCC 2520  
2525 GGAATCGTTTACCAAGCAATGGCATGTACCATCAAAATGCGAGCTGTGTTGGAACATCT 2584  
2521 TTCTGACGAGCTCTGTTTACATTCAGAAAAATGTGTCAATTCCTGCGCCATTTGATTGC 2580  
2585 AGATGTAACAGCAGTGTATACATTGAGGAGGCTGCATCATCCCTGCCCTCAGACTGC 2644  
2581 AGTTAAGGATGTGTAGTTGGGGTCTGAGTGTCTGATTTGTTGAAATGGAGTGAAG 2640  
2645 AAGCTCAGTGTGTGTCAACTGTGCGGCTGCAAGAGTCTCTGTGGAGTGTGTGAAG 2704  
2641 ATTTCGATCCAAATGGCTTAAAGAAAAACCTTACAAATGAGGAGCAGCATGTCCTCAACTG 2700  
2705 GTTCTTCTAATGCTGCTGGAAGAAACCAATATATGAGGAGAGGCTTGGCCCCAACTG 2764  
2701 GATC-----TCAAGAAATCAGGTACATGAGCAGTCCCATGTTACAGTGAAGTCAATCAG 2754

2765 GACCATGTCAACAGGACACAGGTGTATGAGTTGTCTCCATGCCACAGTGAAGTCAACACAG 2824  
2755 TATTCTGGGTTGTGAACACATGGTCTTTCATCAAAAACAAATGAGTGAAGTCCCTG 2814  
2825 TACCTAAGGTCAAGAGCCCTGGAGCATCTCAAGGTGACCTTTGTGAATATGGGGAG 2884  
2815 CCGTGTGGAGGAGAACACAATCTAGGAAAAATCAGATGTGT---GAATACTCGGATGT 2871  
2885 AACTGTGGAGAGGGGTGCAACCCGCAAAAGTGAAGTGCATCAGAAATACAGCAGATGSC 2944  
2872 GAGGTGGAGCAGTGGATAGCAACCTGTGACACAGGATGAATTTCCCCAGAAAAACCCAG 2931  
2945 CCTTCTGAACATGTAGAGATTACCTCTGTGAGCCAGAGAGATGCCCTGGGCTCTAGA 3004  
2932 TCTGTGTCTCTTATGTGTCCCAATGAGTGTGTCTGTCTGAGTGGGAGCTTTTGGAGCAAA 2991  
3005 GTGTGCAAAATTACCATGCCCTGAGGACTGTGTGATATCTGAATGGGTCCATGGACCCAA 3064  
2992 TGCCCCACAGTCAATGCAATCCCAACAAT---GCAGAGAGAACTCGCCACCTGTCTAGA 3048  
3065 TGTGTTTTGCCCTTGCAATCAAAAGCAGTTTCCGGCAAAAGTCAAGTATCCCATCAGACAA 3124  
3049 CCATCACTGAATCTCAAGGACTTGTGTGAAGACTCACAGGTGACGCTTGCCTCTCTGAAT 3108  
3125 CCAGCTGATGAAGGAGATCTTGCCTTAATGCTGTGAGAAAGAACCTGTAACTGTGAC 3184  
3109 GAAATTTGCTTCCAGTTCCAGTACAACTTAACAGAGTGGAGCACAATGCCAGCTGATGAA 3168  
3185 AAAAATGCTGACCTATGATTAATGTAAACAGACTGGAGTACATGTGAGTGTGAG 3244  
3169 AACGCACTGTGTGCAAGGCTCAGGACCCGCTGCTAAGCTGTGTGAGTGTGATGATGSC 3228  
3245 AAGGCAATTTGTGGAATTTGGAATTAACAACAGGATTTGGATTGTGTGAGTGTGATGSC 3304  
3229 AAGCAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 3288  
3305 AAGTCAATTTGACCTGAAATATTTGAAAGCGCTTGGCTTGGAGAAACCTGGCAGATGAAC 3364  
3289 ATTCCCTGCTTGGTGAATGCGTGTCACTGTGCTGAGTGTGAGGCTGAGCGCTTGGACA 3348  
3365 AGTCTCTGATGTTGGAATGCGCTGTGAACTGTGAGTGTGATGTTGCTCTCTTGGTCA 3424  
3349 GAGTGTTCACAGACCTGTGGCAGTGGAGTCAATGAGCCGAGTCAATTTATCAATTATG 3408  
3425 GAATGTTCTCAAACTATGTCGCTCACAGGAAAAATGATCCGAAGACGAACAGTGAAGCCAG 3484  
3409 CCAACCCAGAGGAGGAGCGGCATGCCCAACAGAGCTTACCCAGAGAAACCTGCCCA 3468  
3485 CCTTTCAAGGTGATGGAAGACCATGCCCTTCCCTGATGGAACCTCCAAACCTGCCCA 3544  
3469 GTGACCCCTGCTACAGCTGGGTCTTGGCAACTGGTCTGATGTAAATTTGAGGAGTGA 3528  
3545 GTGAAGCCTTGTATCGGTGGCAATATGGCAGTGTCTCCATGCCAAGTGCAGAGGCC 3604  
3529 GACTGTGGGAGAGGAGTTCAGATCCGAGCCCTTCTGAGTGTGTCAGTGTGTCAGTGTTCATA 3588  
3605 CAGTGTGGAGAGGAGCCAGAACAGGAAACATTTCTGTGTAGTAAATGATGGTGGTCACT 3664  
3589 TCTCATGAGCTGGAGCTGTGAGGATGCACTGTGTGAGAAATGCCCTT----- 3638  
3665 GATGATTTTCAAGCAAGTGTGTGATGAGGAAATCTGTGCTGACATTTGAACATTAATGAT 3724  
3639 -----TCAGGACAGCATCTGGAAGCAGTGTGTCTGTGCTTGGCCCTTGGCCAGGAGACTGCCAT 3693  
3725 GGTAAATAAAATATGTTCTGTGAGGAAATCTGTGAGCCAGCTTGGCCAGGTGACTGTTAT 3784  
3694 TTAAACAGATGTTCAGAGTGGAGCACAATGTAATTAACCTGATTTGAGGAGAGTGT 3753  
3785 TTGAGGACTGTCTTCTGAGGAGCTGTGTGAGTGTGATGAGTGTGAGTGTGAGTGTGAGTGT 3844  
3754 GAGACTGTGGCGCCAGTCTAGATCAAGGACTTTTATAATTCAGTCTTTTGGAGACCAA 3813  
3845 GGTCTTGGTGAATACAGGTGAGTCCAGACCGGTGATTATACAGAACTAGAGAAATCAG 3904

3814	QY	GACAGCTGCCCCCAACAGGTTCTAGAAACACAGCCCTTGTCAGGAGGCAAAATGTTATCAC	3873
3905	Db	CATCTGTGCCAGAGCAGATGTTAGAAACAAATCATGTTATGATGACAGTCTATGAA	3964
3874	QY	TACACATGGAAACCAAGTCCTTTGGAACAATAACGAACGAACACTGATGTTGGTCCAGCGTTCA	3933
3965	Db	TATAATGGATGGCCAGTCGCTTGGAAAGGGCTCTTCCGGAACAGTGTGTGTTCAAGGTCA	4024
3934	QY	GATGGGCTTAATGTCAAGGAGGCTGCTCCCTCAGGCCCGTCTCCTGCTGCCATTTCGGCAG	3993
4025	Db	GATGGTATAAATGTAAACAGGGGGCTGCTTGGTGATGAGCCAGCCTGATGCCGACAGTCT	4084
3994	QY	TGCATTCACGCTGCAGAAAACCTTTCTCTACTGTATACACAGGGTGAGTCTGTGGTTGT	4053
4085	Db	TGTAAACCCCGTGTAGTCAACCCCACTCGTACTGTAGCCAGACAAACCATGCCATTGT	4144
4054	QY	GAGAAGGGCTATACAGAGATAAATGAAATCAAATGGTTTCTCGATTACTGCA-----	4105
4145	Db	GAAGAAGGGTACACTGAAGTCATGTCTTTCTAAACAGCACCTTGAGCAATGCAACATTATC	4204
4106	QY	-----TGAAAGTACCAGGCTCAGAGGATAAAAGCTGATGTGAAAAACCTTTCTG-G	4157
4205	Db	CCCGTGGTGATTAACCCACCATGGAGGACAAAGAGGAGATGTGAAAACCATCGGGCT	4264
4158	QY	GAAAAACAGACCTGTGAAATTCAAAAATACATGATATTTTTAAAGGA--TGGTCTCTTCAA	4215
4265	Db	GTACATCCAAACCAACCTCCACTAACCAGCAGGACGGGAGAGGACCTGGTTCTTACAG	4324
4216	QY	CCACTTGATCCAGATGCCGAGTAAAAAATTTGGTGTATAGCGCTTCAGGTGCGGCTTT	4275
4325	Db	CCATTTGGGCGAGATGGGAGACTAAAGACCTGGGTTTACGGTGTAGCAGCTGGGGCATTT	4384
4276	QY	CTCATCATGATTTTCCTAATATTACTTCTACTCTGTTTGGCAAGAGCA---AAACCA	4332
4385	Db	GTGTTACTCATCTTTATTTGTTCTCCATGATTATCTAGCTTGAAAAGCCAAAGAAACC	4444
4333	QY	CATCAAGCACACCTCCCAACAGAGACCTCTGACCTTAGCCTACGATGAGACTTTAGAC	4392
4445	Db	CAAGAAGGCAAAACCAACCGACTGAAACCTTTAACCTTAGCCTATGATGAGATGCCGAC	4504
4393	QY	ATGTAA 4398	
4505	Db	ATGTAA 4510	
RESULT 6			
LOCUS	AX079870	6373 bp	DNA linear PAT 22-FEB-2001
DEFINITION	Sequence 1 from Patent WO0105971.		
ACCESSION	AX079870		
VERSION	AX079870.1	GI:13159395	
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	Shinkets, R.A. and Fernandes, E.		
TITLE	Secreted polypeptides and corresponding polynucleotides		
JOURNAL	Patent: WO 0105971-A 1 25-JAN-2001;		
FEATURES	Curagen Corporation (US)		
source	Location/Qualifiers		
	1..6373		
	/organism="Homo sapiens"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:9606"		
	235..5001		
	/note="unnamed protein product"		
	/codon_start=1		
	/protein_id="CAC32422.1"		
	/db_xref="GI:13159396"		
	/db_xref="R01575:EMBL:CAC32422"		
CDS			



QY 574 ATGCCCAAGACTGTGAACCTCCAGTGGTCTCTCTGGAGCCCTGTCTCCAGACATGCG 633  
DB |||||  
1096 ATCACAAAGAGTGCAGGTTTCCAGTGGTCCAGAGTGGAGCCCTGTCTCAAAACATGC 1155  
QY 634 CGTTCCAGGAGTCTCTTGGCAGGATTTAGGAGCAGGAGCCGGAACGTCAAGCACATGGCT 693  
DB |||||  
1156 CATGACATGGTCCCTCCAGGCACTCGTAAAGCACAGAACCATCAGCAGTTTCCC 1215  
QY 694 ATTGAGGTGGAAGAGTGTCTCTGAATCTTTGAGAAAGAGGCTGTCAATTTGTAAGGA 753  
DB |||||  
1216 ATTGGCAGTGAAGAGAGTGTCCAGATTTGAAGAAAAAGAACCTGTCTCAAGGA 1275  
QY 754 GAACTTCTGCACGAATGTCACAGGTATCTCTGGAGAACTTCTGAATGGAAGAAATGCCAA 813  
DB |||||  
1276 GATGGAGTTGTCCTGTGCCACGTATGGCTGGAGAACTACAGAGTGAAGTGTGCGGT 1335  
QY 814 GTCCTCTCTCTCTCGAGCAGCAGGATCCCCACTGGCATGTGAACGGGACCCGTGTGGC 873  
DB |||||  
1336 GTGGACCCCTTTGCTCAGTCAAGCAGCAAGAGGCGCGCAACAGACGGCCCTCTGTGGA 1395  
QY 874 GGTGGGATCCAGACCCGGAGGTGTACTGTGCCAG-----AGGTACCAGCA 921  
DB |||||  
1396 GGGGATCCAGACCCGAGAGGTGTACTGTGGTGGAGGCAACGAAACCTCTCTCAAA 1455  
QY 922 GCTGCCGCACTGAGGSCCAAGGAAGTCTCTAGACCTGTGGAAGGCAATTAATGTGGGA 981  
DB |||||  
1456 TTAAGTACCACAAGAAACAAAGAGCCCTCAAAGCCATGGACTTAATAATATGCAC'TGA 1515  
QY 982 CCCGCCGCTTGCCTCTCAGTCTGCAATATCCCTTGTCTACGGACTGCATAGTATCT 1041  
DB |||||  
1516 CTAATCCCTTAATACACAGCTGTGGCACAATCCCTTGTCCAACCTGAATGTGAAGTTTCA 1575  
QY 1042 TCTCGTCAAGCTGGGCTGTGCATCATGAAACTGTGTAACCTCAGGCGGAAAAA 1101  
DB |||||  
1576 CTTTGGTCAAGCTGGGACCTTGTATTAAGAACTGTATGATCAGCAGGGAAGAAA 1635  
QY 1102 GGAATTAGAAAGAGCAGCG-----CAATGTCCTCATGGAATCTACAGGCGCTGCA 1152  
DB |||||  
1636 GGCCTTCAAACTGAGGAAGCGGCAATTAACCAATGAGCCCACTGGAGGCTCTGGGGTAAAC 1695  
QY 1153 GGGCATTCGCCCTCATTTGGTGGAGTCTGTCTCTGTGAGGATCCAAATGTGTACCGATGG 1212  
DB |||||  
1696 GGAATCGCCCTCACTTACTGGAAGCAATCCCTGTGAAGAGCCTGCTGTATGACTGG 1755  
QY 1213 ---CTGGCATCAGAAGGATCTGTTCCCTGATCATGGAA---AATGTGSCCTGGACAT 1266  
DB |||||  
1756 AAAGCGGTGAGACTGGGAGACTGCGAGCCAGATAACGGAAGGAGTGTGTCCAGGCACG 1815  
QY 1267 CGTATTTCAAGGCGGTCTGCCAGATGACCGCGGAGAGATGTATCAGGAGTCTTTGC 1326  
DB |||||  
1816 CAAAGTTCAAGAGGTGTGTGCATCAACAGTGAAGGAAGAGTTGACAGACAGCTGTGC 1875  
QY 1327 CCAGTTCCCTCTCTCTGAGAGGAAGTCTTGTGAAATTTCCCTGCCGAATGGAGTGTGTG 1386  
DB |||||  
1876 AGAGATGCCATCTTCCCATCCCTGTGSCCTGTGATGCCCATGCGCGAAGACTGTGTG 1395  
QY 1387 CTGAGCGAGTGAACGAGTGTCTATCTGTCCAGTCTCTGTTCAAATTAATAAATCTCAGAT 1446  
DB |||||  
1936 CTCAGCAATGCTCTACGTGTCTCTCTGCTCACAACCTCTCTCAGGGAAGAACAGACGAA 1995  
QY 1447 GGGAAACAGACAGCAGGTCAAGAACTATCTCTGGG---ACTGGCTGGGGAAGGTGGAAAGCCA 1503  
DB |||||  
1996 GGGNAACAGATACGAGCAGATCCATCTCTGGCCCTATGCGGGTGAAGAGTGGAAATTCGC 2055  
QY 1504 TGTCCCTCTAGTCAGGTCTCCAGAGCATCGTTTGTGTAAAGCAATTCCTGTATGACAG 1563  
DB |||||  
2056 TGTCCAAATAGCAGTGTCTTTCAGAGATGACGAAGCTGTAAATGAGCATCTCTTGCACAGTG 2115  
QY 1564 CTTCACTGGGAGACATCGCTTGGGCGCTTGTCTGAGGACATTTGGTAACTGGCCCTT 1623  
DB |||||  
2116 TACCACTGGCAACTGGTCTCTGGGCGCAGTGCATTTGAGGACACCTCAGTATCTGCTTTC 2175

QY 1624 AATGCAACCAATTTGGCTGGGAATGGAGAGCCACGTTGTGTGTAGGCAATTCAGACTCGGAGA 1683  
DB |||||  
2176 AACACAACACTACGACTTTGGAAATGGGAGGCGCTCTGTCTCTGTGGCATGCAGACAAGAAA 2235  
QY 1684 GTCTTCTGTCTCAAGAGTCAAGTGGGCAAGTAATGACCAAAAGAGATGTCAGATTTCTACT 1743  
DB |||||  
2236 GTCATCTGTGTGGAGTCAATGTGGGCCAAGTGGGACCCAAAATGTCTCTGAAGCCCTT 2295  
QY 1744 CGACTGTAAACTGTGGGCCCTGTCTTCTCCATGCAAAAAAGACTGTATTTGTGACTGTCT 1803  
DB |||||  
2296 CGACCTGAAACTGTAGGCGCTGTCTGTCTCTTGTAAAGAGGACTGTATTTGTGACCCCA 2355  
QY 1804 TTCAGTGTGTGACACCTCGCCCAAGGATGTCAGCAGCAGGAATGCGCACACT---AABA 1860  
DB |||||  
2356 TATAGTGTGATGACATCATGCCCCCTTCTGTGTAAAGAGGGGACTCCAGTATCAGGAAG 2415  
QY 1861 CAGTCTCGATACAGAAATCATATCCAAAGAGCAGCAATGGAGGCCAGGAATGCCCAGAT 1920  
DB |||||  
2416 CAGTCTAGGCATCGGCTCATCATTCAGCTGCCAGCAACCGGGGCGGAGACTGCACAGAT 2475  
QY 1921 ACCTTATATGAGGAGAGAGTGTGAAGATGTTTCTGTGTCTCTGTATATCGGTGGAG 1980  
DB |||||  
2476 CCCCCTTATGAAGAGAGGCGCTGTGAGGCACCTCAAGCGTGCCAAAGCTTACAGGTGGAAG 2535  
QY 1981 CCACAGAAATGGAGCCCTTGCATCTTAGTCCAGAGTCTGTCTGGCAGGGAATAACGGGC 2040  
DB |||||  
2536 ACTCACAAATGGCGCAGATGCCAATTAGTCCCTTGAGCGGTGCAACAGACAGCCCTGGA 2595  
QY 2041 AGCAGTGAAGCCCTGTGGAAAGGGTTTCAAAACAGAGTGTCTCATGCAATCTCTGATGAC 2100  
DB |||||  
2596 GCACAGMAAGGCTGTGGGCTGTGGCGACAGGCAAGAGCCATTTACTTGTGCAAGCAAGAT 2655  
QY 2101 AACCGTCAAGCAAAATGATGAAATGCTCAAGCAGACAAAACGGCATGCTCTCTCTTGTG 2160  
DB |||||  
2656 GGAGGACAGCTGGAAATCCATGAGTGCCTACAGTATGACGGCCCTGTGCCAGCCCTTACC 2715  
QY 2161 CAAGAAATGACAGTCCCATGTTCGAGAGACTGCACTTCACTGTCTGTGGTTCGAAGTTAGC 2220  
DB |||||  
2716 CAGGCTTGCAGATCCCTTGCAGGAGTACTGTCAATTGACCAAGCTGTGTCGAAGTTTCT 2775  
QY 2221 CCTGTCTCCACGAATGTGAAGCCACAAAAGTAGCGCGCACACTCACAGGGAAGAC 2280  
DB |||||  
2776 TCAATGCAATGGAGACTGTGTGTGAGTGGACCAAGAGCGCACTCTTGTGTGAAAGAT 2835  
QY 2281 AGAAGAAGGAGAAATGCCAGGATTTGACCTTTACCCTCTAGTGGAGACAGAACTATGT 2340  
DB |||||  
2836 AAAAGAAGGAAATGTAAAAATTTCCCATTTGTATCCCTGATTTGAGACTCAGTATTTGT 2895  
QY 2341 CCTTGTGATGAATTTATATCCCAACCTTATGGAACCTGGTCAGATTCATTTCTTCAGAA 2400  
DB |||||  
2896 CCTTGTGCAAAATATAATGCAACCTGTGGGGAATGGTCAGACTGTATTTTACCAAG 2955  
QY 2401 GGCAGAAAGGAGCCTCACCGAGGACTGGGGGTACAAGCAGACAGCAAAAGAAATGTGGAGAA 2460  
DB |||||  
2956 GGAAGAGTGAAGTGTGTCTGGAAATGAAGTACAGGAGACATCAAGGAATGCGACAA 3015  
QY 2461 GGCCTCGCTTTTCGAGCAGTAGCCTGTCTGTATAAAATGGAAGACCTGTTGACCCCTTCC 2520  
DB |||||  
3016 GGATATCGTTTACCAAGCAATGGCATGTACGATCAAAATGGCAGGCTTGTGGAACATCT 3075  
QY 2521 TTTCTGACAGCAGCTGTGGTTTACATTTCAAGAAAAATGTGTCATTTCCCTGCCCATTTTCAATTGC 2580  
DB |||||  
3076 AGATGTAAACAGCCATGGTTTACATTTGAGGAGGCTGCAATCCCTGCCCCCTCAGACTGC 3135  
QY 2581 AGTTTAAGCAATGGTCTAGTTGGGGGCTTTGCAATTCATTTGTGGAATTTGGAAGTGAGA 2640  
DB |||||  
3136 AAGCTCAGTGAAGTGTCTCAACTGTGTGCGGCTGTGCAAGAGTCTGTGTGGAGTGGTGTGAAG 3195  
QY 2641 ATTTCGAATCCAAATGCTAAAGAAAACTTACAAATGGAGGACGACCATGTCCCAAACTG 2700  
DB |||||  
3196 GTTCGTTCTTAATGGCTGGTGAAGAAACCATATAATGAGGAAGGCGCTTGTGCCCAAACTG 3255  
QY 2701 GATC-----TCAAGAATCAGGTATACAGGCGAGTCCCATGTTTACAGTGAAGTCAATCAG 2754

```

3256 GACCATGTCACAGGACAGAGTGTATGAGGTTGTCCTCATGCCAGTGAACACAG 3315
3275 TATTCCTGGGTTGTAGAAACACATGCTTCTCATGCAAAATCAACAATGAGCTGAGGTCCTTG 2814
3316 TACCTATGGGTACAGAGCCCTGGAGCATCTGCAAGGTGACCTTTGTGAATATGCGGAG 3375
2815 CGCTGTGGAGGAGAACACAATCTAGAAATACAGATGTGT---GAATATCTCGGATGT 2871
3376 AACTGTGGAGAGGGCTGCAAAACCCGAAAGTGAATGATGATGATGATGATGATGATGATGAT 3435
2872 GAAGTGGAGCAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2931
3436 CCTTCTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3495
2932 TCTGTGTTCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2991
3496 GTGTGCAAAATACCAATGCTGAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 3555
2992 TGCCACAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3048
3556 TGTGTTTGGCTTGGCAATCAAAAGCAGTTTCCGCAAGGTGATGATGATGATGATGATGATGAT 3615
3049 CCATCACTGAACTCAAGGACTTGTGCTGAAGACTCACAGGTGAGCTTGCCTCTCTGAT 3108
3616 CCAGCTGATGAAGAGATCTTGGCTTATGCTGTGTAAGAGAACCTGTGAACCTGATGAC 3675
3109 GAAATTTGCTTCAGTTTCAGTATCAATCTAACAGAGTGAGCACATGCCAGCTGATGAA 3168
3676 AAAAACTGCTTACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3735
3169 AAGCCACCTGTGTTCAAGGCTGAGGACCGCTGCTAAGCTGTGTTGAGTGTGTTGAGTGTGATG 3228
3736 AAGCCAGTTTGTGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAAT 3795
3229 AAGCCAGTCAAGATGACCAATGTGAGAGCAATAATTTGGAGAGCCCGCAGAGATGAGC 3288
3796 AAGTCAGTTGACCTGAATATTTGGAAGGCTTGGCTTGGAGAGAACTGGCAGATGAGC 3855
3289 ATTCCCTGCTTGGTGAATGCTGTGCTCACTGCTCACTGCTCACTGCTCACTGCTCACTGCTCA 3348
3856 ACGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3915
3349 GAGTGTTCAGACCTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3408
3916 GAATGTTCTCAACATGTGCTTACAGGAAATGATGATGATGATGATGATGATGATGATGATGAT 3975
3409 CCACCCAGGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3468
3976 CCTTTTCAAGGTGATGGAAGAGCAGTCCCTTCCCTGATGAGCAGTCCAAACCCCTGCCCCA 4035
3469 GTGACCCCTGCTACAGCTGGGTCCTTGGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3528
4036 GTGAAGCTTGTATCGTGGGCAATATGAGCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4095
3529 GACTGTGGGAGAGGATTCAGATCCGACCTTCTGATGATGATGATGATGATGATGATGATGATGAT 3588
4096 CAGTGTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4155
3589 TCTCATGAGCTGAGCTGCTGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3638
4156 GATGATTTGAGCAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4215
3639 -----TCAGGACAGCATCTGAGCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3693
4216 GGTAAATAAAATATGTTCTGAGGAGTCTGCTGAGGAGTCTGCTGAGGAGTCTGCTGAGGAGT 4275
3694 TTAACAGATGCTCAGAGTGGAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3753
4276 TTGAAGACTGTGCTTCTGAGGAGCTGCTGCTGAGCAGTGTGCTGAGTGTGATGATGATGATGAT 4335
3754 GAGACTGTGGGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3813
```

```

4336 GGCTTTGGTGGAAATACAGGTTCAGATCCAGACCGGTGATTATATCAAGAACTAGAGATCAG 4395
3814 GACAGCTGCCGCCCAACAGGTTCTAGAAACAGCGCTTGTACAGAGGCAATGTTATCAC 3873
4396 CATCTGTGCCCAGAGCAGATGTTAGAAACAAATCATGTTATGATGACAGTGTCTATGAA 4455
3874 TACACATGGAAGCAAGCTCTTTGGAAACAATAAGCAACGAACCTGTATGTCGCCAGCGTTCA 3933
4456 TATAATGGATGGCCAGTGTCTTGGAGGGCTCTTCCGAAACAGTGTGTCTCAAGGTCA 4515
3934 GATGCGCTTAATGTACAGAGGCTGCTCCCTCAGCGCCGCTCTGCTGTCATTCGGCAG 3993
4516 GATGGTATTAATGTAAACAGGGGCTGCTTGGTATGAGCCAGCTGATGCGCAGCTCT 4575
3994 TGCAATTCAGCTGACAGAAACCTTTCTCTACTGTACACAGGCTGGAGTCTGTGGTGTGT 4053
4576 TGTAAACCCACCGTGTAGTCAACCCACTCGTACTGTAGCGAGACAAAACATGCCATGT 4635
4054 GAGAAGGCTTATCAGAGATAATGAATCAATGATGATGATGATGATGATGATGATGATGATGATGAT 4105
4636 GAAGAGGGTACACTGAAGTCTGTTCTTACAGAGCACCTTGGCAATGCACACTTATC 4695
4106 -----TGAAGTACCAGGCTCAGAGGATAAAAGCTGATGTGAAAACCTTTCTG-G 4157
4696 CCCGTGTGTGTTTACCCACATGGAGGACAAAGAGAGATGTGAAAACCTGCGGCT 4755
4158 GAAAAACAGACCTGTGATTTCAAAATATACATGATATTTTAAAGGA--TGGTCTCTTCAA 4215
4756 GTACATCAACCCAAACCTCCAGTAAACCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4815
4216 CCCTTGTATCCAGATGCGGAGTAAATTTGGGTTTATGGCGTTTCAGGTGGCGCTTTT 4275
4816 CCATTTGGGCGAGTGGAGACTTAAAGACCTTGGGTTTACCGTGTAGCAGCTGGGGCATTT 4875
4276 CTCACTCATGATTTTCTTAATATTTTACTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 4332
4876 GTGTTACTCATCTTTATTTGTTCTCATGATTTTATCTAGCTTGCAGAAAGCCCAAGAAACCC 4395
4333 CATCAAGACACCTCCCAACAGAGGCTCTGACCTTACCTTACCTTACCTTACCTTACCTTACCT 4392
4936 CAAAGAGGCAAAACACACCGACTGAAACCTTTAACTTACCTTACCTTACCTTACCTTACCT 4995
4393 ATGTAA 4398
4996 ATGTAA 5001

RESULT 7
AR338841
LOCUS AR338841 6378 bp DNA linear PAT 17-AUG-2003
DEFINITION Sequence 332 from patent US 6569662.
ACCESSION AR338841
VERSION AR338841.1 GI:33725698
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 6378)
AUTHORS Tang,Y.T., Zhou,P. and Dmanac,R.T.
TITLE Nucleic acids and polypeptides
JOURNAL Patent: US 6569662-A 332 27-MAY-2003;
FEATURES Location/Qualifiers
source 1..6378
/mol_type="genomic DNA"

Query Match 28.0%; Score 1229.6; DB 6; Length 6378;
Best Local Similarity 57.6%; Pred. No. 0;
Matches 2596; Conservative 0; Mismatches 1784; Indels 126; Gaps 17;

QY 10 TGCAATTCAGAGCTGAACCGAGCTGTGTTGCAATGAAATATGCAACACTTGGCCTT 69
```



Db 541 TGCATCCAGAAAGCAAAAGACATTCCTCGGAGGATATCATCTGTGAGTACTTTGAGCCC 600  
Qy 70 CAGCCTCCTACAGAACAGGCTTGCTCATCTCTTGTCCCGGATTTGTGTAGTATCTGAG 129  
Db 601 AAGCCTCTCCCTGGAGACAGGCTTGCTCATCTCTTGTCCAGCAAGATTGCATCGTGTCTGAA 660  
Qy 130 TTCTTACCATGCTCCAACTGTAGCAAGGATGTGGAGAAATGTCAGCATAGACTCGC 189  
Db 661 TTTTCTGCTGCTCGAAATGCTCCAAAGACTTCCGAGCGGCTTCCAGCACCGAGCGCT 720  
Qy 190 CGGTCATAGTCCCTCTCTTTGGTGGTTTGAATGTCCAAATCTGACTGATCAAGA 249  
Db 721 CATGTGGTGGCGCCCGCCAGTTCGGAGGCTCTGGCTGTCCAACTGACGGAGTTCCAG 780  
Qy 250 GCCTGTGATGCTCCCATTTCTGTCTCTTGGGAAGAGAAATATATATTTAGCTTTAG 309  
Db 781 GTGTG-----CCAACTCAGTCCATGCGAGCCGAGGCTTCAGGTACAGCCTGCAT 831  
Qy 310 GTTGGACCATGGAGTAATGACAGACTGCTCATCTTAAAGAAATTAATCCAAGCGGAAGA 369  
Db 832 GTGGGCCCTGGAGACCTGCTCATGTGCCCACTCCCGACAGTAAAGCAAGCAAGGAGA 891  
Qy 370 ACTGTTCTGGATTTTAA-----CTCTGATTCAAAATGAG 402  
Db 892 CGCGGGAAGATAAAGAACGGAAGGACCGCAGCAAGAGGATTAAGGATCCAGAGCC 951  
Qy 403 CGAGTCACTTTAAACATCAAGTTACA-----AAGCACATCATCTTCCAGTCT 453  
Db 952 CGCAGCTTATTAGAAAGAGAAACAGAAACAGGACAGACAGCAAGAGAACAAATAT 1011  
Qy 454 TGGCAATPAGAGATAGTATTCAAAACCGGAGGTTTCTGTACAGAAAGTGTAGCAAA 513  
Db 1012 TGGACATCCAGATTGGATATCAGACAGAGAGGTTATGTGCTAACAAGACGGGAAA 1071  
Qy 514 AATGCTATGTTAAGCCTTTGCTTCAAGATTCCTTCCATGTAGTGTTCAGTCTGCTGATC 573  
Db 1072 GCTGCTGATTTAAGCTTTTGCACAGAGAGAGCTTCCAAATGACCTTCCAGTCTCTGTGTG 1131  
Qy 574 ATGCCCAAAGACTGTGAACCTCCAGTGTCTCTCGAGGCCCTGCTCCAAAGACATGC 633  
Db 1132 ATCCAAAGAGTCCAGGTTTCCGAGTGTGTAGAGTGGAGCCCTGCTCAAAAACATGC 1191  
Qy 634 CTTTCAGGAGTCTCTTGCAGAGATTAGAGAGAGGACCGGAACGTGAAGCAGATGGCT 693  
Db 1192 CATGACATGGTGTCCCTGCAAGGACCTCGTGTAAAGGACAGCAACATCAGGAGCTTCCC 1251  
Qy 694 ATTGGAGTGAAGAGGAGTGTCTGAACTTCTCAGAAAGAGGCTGCTATTGTGAGGA 753  
Db 1252 ATTGGAGTGAAGAGGAGTGTCCAGAAATTTGAAGAAAGAGAACCTGTTGTCTCAGGA 1311  
Qy 754 GAACTTTGACAGCAATGTCCAGGATATCTCTGGAGAACTTCTGAATGGAAGAAATGCCAA 813  
Db 1312 GATGGAGTGTCCCTGTGCCACGATATGGCTGGAGAACTACAGAGTGGACTGAGTGGCGT 1371  
Qy 814 GTCTCTCTCTCTCGAGCAGAGGATNCCCACTGGCATGTGAGGACCCGCTGTGTGGC 873  
Db 1372 GTGACCTTTGCTGTAGTCAGCAGCAAGAGGCGGCAACAGAGCGGCTCTCTGTGGA 1431  
Qy 874 GGTGGGATCCAGACCGGAGGAGTGTACTGTGCCAG-----AGCGTACCAGCA 921  
Db 1432 GGGGCATCCAGACCGAGAGGTTACTGTGCGTGCAGGCCAACGAAACCTCTCTCAAA 1491  
Qy 922 GCTGCCGACTGAGGGCCAGAGAGTCTCTAGACCTGTGGAAGAGGCAATATGTGTGGGA 981  
Db 1492 TTAAGTACCCCAAGAAACAAAGAGGCTCAAGAGCAATGGAACCTTAAATATATGCACTGGA 1551  
Qy 982 CCCGCCCGTGTGCCCTCTCACTCTGCAATATCCCTTGTCTAGGACTGCAATGATATCT 1041  
Db 1552 CCTATCCCTAATATACACAGCTGTGCCAATCTCTGTCCAATGATGTGAAGTTTCA 1611  
Qy 1042 TCCTGGTCAGCTGGGCTGTGCATCCATGAATCTGTCATGAACCTCAGGGGAAAAA 1101  
Db 1612 CTTTGGTCAGCTGGGACCTTGTACTTATGAAACTGTATGATGATCAGCAAGGAAAAA 1671

Qy 1102 GGATTTAGAACGAGCAGCG-----CCATGTCTCATGGAATCTACAGGCGCTGCA 1152  
Db 1672 GGCCTTCAAACTGAGGAAGCGGCGCATTTACCAATGAGCCCACTGGAGGCTCTGGGGTAACC 1731  
Qy 1153 GGGCATTCGCCCTCANTTGGTGAGTCTCTTCTTGTGAGGATCCAAATGTCTACCGATGG 1212  
Db 1732 GGAATCTGCCCTCTACTTACTGGAAGCCATTCCTGTGAAGAGCTGCTCTTATGACTGG 1791  
Qy 1213 CTGGCA---TCAGAAGGATCTGTTTCCCTGATCATGAA---AATGTGGCTGGGACAT 1266  
Db 1792 AAAGCAGTGAAGCTGGAGACTGCGAGCCAGATAACGGAAGAGGTGTGGTCCAGGCAG 1851  
Qy 1267 CGTATTCTGAAGCCGCTCTGCGAATGACCGGAGAGATGTATCAGGAGTCTTTGC 1326  
Db 1852 CAAGTTCAGAGGTTGTGTGCATCAACAGTGTGAGAGAAAGTTGACAGACAGCTGTGC 1911  
Qy 1327 CCAGTTTCCCTCTCTGAGAGGAAAGTCTTGTGAATTCCTGCGGAATGGACTGTGTG 1386  
Db 1912 AGAGATGCCATCTTCCCATCCCTGTGSCCTGTGATGCCCATGCCCGAAGACTGTGTG 1971  
Qy 1387 CTGACGAGTGAACGAGTGTGTCTTCTTCCAGTCTCTTCAAAATAAAAACTCAGAT 1446  
Db 1972 CTCAGCACATGCTTACGTTGCTCTCTCTCTCTCAGACACTCTCAGGGAACACAGAAA 2031  
Qy 1447 GGGAAACAGACCAGTCAAGAACTATCTCTGSC-----ACTGGCTGGGAGGTGGAAAGCCA 1503  
Db 2032 GGGAAACAGATACGAGCAGATCTCTTCTGCGCTATGCGGTGAAGAGGTGGAAATTCG 2091  
Qy 1504 TGTCTCTCTAGTCTCAGGCTCTCCAAGAGCATCGTTTGTGTATATGACCATTCCTGTATGCG 1563  
Db 2092 TGTCTCAATAGCAGTCTTTCAGAAAGTACGAAGCTGTATATGAGTCTCTTGCACAGTG 2151  
Qy 1564 CTTCTCTGGAGACATCGCTTGGGCGCTTGTCTGAGGACATCTGTAATCTGACTGCCCTT 1623  
Db 2152 TACCACTGGCAAACTGGTCTCTGCGGCGAGTGCATGTAGGACACTCAGTATCGTCTTC 2211  
Qy 1624 AATGCAACCATTTGGCTGGAATGGAGAACCCAGCTGTGTGTAGGCAATTCAGACTCGGAGA 1683  
Db 2212 AACACACTAGCATTGGAATGGGAGGCTCTCTGCTGTGCGCATGCACAGCAAGAAA 2271  
Qy 1684 GTCTCTGTGTCAAGAGTCACTGTGGGACAAATATGACCAAAAGATGTCCAGATTTCTACT 1743  
Db 2272 GTCATCTGTGTGCGAGTCAATGTGGGCGCAAGTGGGACCCAAAAAATGTCTCTGAAAGCCTT 2331  
Qy 1744 CGACTGTGAACTGTGCGCCCTGTTTCTCCCATGCAAAAAAGAGCTGTATTGTGACTGCT 1803  
Db 2332 CGACTGTGAACTGTGAGGCTTGTCTGCTTCTTGTGAAGAGGACTGTATTGTGACCCCA 2391  
Qy 1804 TTCAGTGTGACACCTGCTGCCAAGGATGTGCCAAGCAGGAAATGCCACAGT---AAAA 1860  
Db 2392 TATAGTGACTGGACATCATGCGCCCTCTTCGTGTAAAGAGGGGACTCCAGTATCAGGAAG 2451  
Qy 1861 CAGTCTCGATACAGAACTCATCCAGAACGCAATGAGGCGCAGGATGCCAGATGCCAGAT 1920  
Db 2452 CAGTCTAGGCTATCGGCTCATCTTCTGCTGCTGCTTCTGCTGCTATATCGGTGGAAG 2511  
Qy 1921 ACCTTATATGAGGAGAGAGTGTGAAGATGTTCTTGTGTCTGCTATATCGGTGGAAG 1980  
Db 2512 CCCCTCTATGAAGAGAGGCTGTGAGGCACTCAAGCGTCCAAAGCTACAGGTGGAAG 2571  
Qy 1981 CCACAGAAATGAGCCCTTGCAATCTTAGTGCAGAGTCTGTCTGGCAGGGAATACCGGC 2040  
Db 2572 ACTCAAAATGGCGCAGATGCCAATTAGTCTCTTGGAGCGTGCACACAGACAGCCCTGGA 2631  
Qy 2041 AGCAGTGAAGCTGTGGAAAGGGGTTACAAACAGAGCTGTCTCATGCTCTCTGATGAC 2100  
Db 2632 GCACAGAGAGGCTGTGGGCTTGGGCGACAGGAGAGGCAATTTCTTGTGCGAAGCAGAT 2691  
Qy 2101 AACCGGTGACAGAAATGATGAAATGCTCAAGCAGACAAACGGCATGCCCTCTCTTGTG 2160  
Db 2692 GGAGACAGGCTGGAATCCATGAGTGCCTACAGTATGAGGCGCTGTGCCAGCCCTTACC 2751

QY 2161 CAGAAATGCACAGTCCCATGTCGAGAGACTGCACCTTCACTGCTGGTCCCAAGTTTACG 2220  
Db 2752 CAGCCCTGCCAGATCCCTCGCCAGATGATGTCATTTGACCAAGCTGGTCCAAAGTTTTCT 2811  
QY 2221 CCCTGCTCCACGAACTGTGAAGCCACAAAAAGTAGGGCGGCGACAGCTCACAGAGAAAAGC 2280  
Db 2812 TCATGCAATGGAGACTGTGTGTGACGTTAGGACACAGAAAGCGCACTCTTGTGGAAAAGT 2871  
QY 2281 AGAAGAGAGAGAAATGCCAGGATTCGACCTTTACCTCTAGTGGAGACAGAACTATGT 2340  
Db 2872 AAAAAAAAAAAAAATGTAAAAATTCCTATTTGTATCCCTGATTGAGACTCAGTATTGT 2931  
QY 2341 CCCTGTGATGAATTTATATCCCAACCTTTATGAAAACCTGGTCAGATTGCATTTCTCCAGAA 2400  
Db 2932 CCTTGTGACAAATATATGACAACTGTGGGNACTGGTCAGACTGTATTTTACCAGAG 2991  
QY 2401 GGCAGAGGAGCCCTCACCGAGGACTCGGGTACAAGCAGACAGCAAGAAATGTGAGAA 2460  
Db 2992 GGAAGAGTGAAGTGTCTGGGAAATGAAAGTACAAGGAGACATCAAGGAATCGGACAA 3051  
QY 2461 GGCTCGCTTTCGAGCAGTAGCTGTTCTGTATAAAATGGAAGACCTGTGTGACCCCTCC 2520  
Db 3052 GGATATCGTTTACCAGCAATGGCATGCTACGATCAAAATGGCAGGCTTGTGGAACATCT 3111  
QY 2521 TTCTGCAGCAGCTCTGCTTACATTCAGAAAAAATGTGTCAATTCCTGCCCAATTTGATTGC 2580  
Db 3112 AGATGTAACAGCCATGTTACATTTAGAGGCGCTGCATCATCCCTGCCCTCAGACTGC 3171  
QY 2581 AGTTAAGCAATGGTCTAGTGGGGTCTTGGAGTTTCATCTTGTGGAATGGAGTGAGA 2640  
Db 3172 AAGCTCAGTGAAGTGTCTCACTGGTCCGCTGAGCAAGTCCCTGTGGGAGTGTGTGAAG 3231  
QY 2641 ATTCGATCCAAATGGCTTAAAGAAAAACCTTACAATGGAGGAGACCATGTCGCCAACTG 2700  
Db 3232 GTTCGTTCTAAATGGCTGCGTGAATAAACCAATATATGGAGAAAGGCTTGCCCAAACTG 3291  
QY 2701 GATC-----TCAGAAATCAGGTACATGAGGCACTGCCATGTTACAGTGAAGTCAATCAG 2754  
Db 3292 GACCATGTCAACAGGACAGGTGTATGAGTGTGCCATGCCACAGTGACTGCAACCG 3351  
QY 2755 TATTCCTGGCTGTAGAACACTGGTCTTCAATGAAATCAACAATGAGCTGAGTGCCTCG 2814  
Db 3352 TACCTATGGGTACAGAGCCCTGGAGCATCTGCAAGGTGACCTTTGTGAATATGCGGAG 3411  
QY 2815 CGCTGTGGAGGAGAACACAATCTAGAAAAATCAGATGTCT---GAATACTGCGGATGT 2871  
Db 3412 AACTGTGGAGAGGCGTGCAACCCGAAAGTGAATGATGCATGCAGATACAGCAGATGSC 3471  
QY 2872 GAAGGTGGAGCAGTGGATAGCAACCTGTGCAACAGAGATGAATTCGCCAGAAACCCAG 2931  
Db 3472 CCTTCTGAACATGTAGAGGATTACCTCTGTGACCCAGAGAGATGCCCCCTGGGCTCTAG 3531  
QY 2932 TCCTGTTCTCTTATGTGTCCTCAATGATGTGTCATGCTGAGTGGGAGCTTTTGGAGCAA 2991  
Db 3532 GTGTGCAAAATACATGCCCTGAGGACTGTGTGATATCTGAATGGGGTCCATGGACCCAA 3591  
QY 2992 TGCCCAAGCTATGCGATCCCAACAAT---GCAGAGAGAACTGCCACCTGTCTAAGA 3048  
Db 3592 TGTGTTTGGCTTGGCAATCAAGACGTTTCCGCAAAAGGTGACCTGATCCCATCAGACAA 3651  
QY 3049 CCATCACTGAATCAAGGACTGTGTGTAAGACTCACAGTGGAGCTTGCCTCTCTGAAT 3108  
Db 3652 CCAGCTGATGAAGGAAGATCTCCCTTAATGTCTTGTGAGAAAGAACCTGTAACTCTGAAC 3711  
QY 3109 GAAAAATGCTTCCAGTTCAGTCAATCTAACAGAGTGGAGCACATGCCAGCTGAGTGAA 3168  
Db 3712 AAAAACTGTACCACTATGATTAATATGTAACAGACTGAGATACATGTCACTGAGTGAG 3771  
QY 3169 AACGCACTCTGTGGTCAAGCGCTCAGGACCCGCTGCTAAGCTGTGTGTGCAAGTATGCC 3228  
Db 3772 AAGGCAGTTTGTGGAATGGAATGAAATAAAAAAAGGATGTGTGATTTGTGTTGCAAGTATGCC 3831  
QY 3229 AAGCCAGTCAGCATGGACCAATGTGAGCAGCAATAATTTGGAGAGGCCCCCAGAGATGAGC 3288

Db 3832 AAGTCAGTTGACCTGAAATATTTGTGAAGCGCTTGGCTTTGGAGAAAGAACTGGCAGATGAAC 3891  
QY 3289 ATTCCCTGCTTGGTGGAAATGGTGTCAACTGTAGCTCTCAGGCTTGGAGCGGCTTGGACA 3348  
Db 3892 AGCTCTCGATGTTGGAATGCCCTGTGAACCTGTAGCTTCTGATTGCTCTCTTGGTCA 3951  
QY 3349 GAGTGTTCACAGACCTGTGGCCATGGAGGTGGAATGAGCCGAGACTCGATTATCATTTAG 3408  
Db 3952 GAATGTTCTCAACATGTGGCTTCACAGAAAAATGATCCGAGACGAAACAGTGACCCAG 4011  
QY 3409 CCAACCCCAAGAGAGAGGAGCGCCATGCCACAGAGCTTACCCAGAGAAACCTGGCCCA 3468  
Db 4012 CCCTTTCAAGGTGATGGAGACCAATGCCCTTCCCTGATGGACCACTCCAAACCTTGCCCA 4071  
QY 3469 GTGACCCCTGCTACAGCTGGGTCTTGGCAACTGGTTCGATGTAAATTTGGAGGGTGA 3528  
Db 4072 GTGAAGCCCTGTTATCGGTGGCAATATGGCCAGTGGTCTCCATGCCAAGTCAGAGGCC 4131  
QY 3529 GACTGTGGGGAAGGAGTTTCAGATCCGAGCCCTTCCCTGCACTGGTCCACAGTGGTTCATA 3588  
Db 4132 CAGTGTGGAGAGGAGCCAGAAACAGGAACATTTCTTGTGTAGTAAGTATGGGTGAGCT 4191  
QY 3589 TCTCATGCACTGGACGCTGTCAGATGCACTGTGTGAGAAATGCCCTT----- 3638  
Db 4192 GATGATTTCAACAAGTGTGTGATGAGGAATTTCTGTCTGACATTTGAACCTATTATAGAT 4251  
QY 3639 -----TCAGGACAGCATCCTGAAGAGCTGTGTCTGTGCCCTGCCAGGAGACTGCCAT 3693  
Db 4252 GGTATAAAAAATATGGTTCGTGGAGGAATCCTGACAGCCAGCCCTGCCAGGTGACTGTAT 4311  
QY 3694 TTACAGAAATGCTCAGAGTGGAGCAGATGTGAATTAACCTGCATTGATGGAAGAAGCTTT 3753  
Db 4312 TTGAAGGACTGCTTCTCTGGAGCCTGTGTGAGCTGACCTGTGTGAATGTGTGAGATCTA 4371  
QY 3754 GAGACTGTGGGCCGCCAAGCTTAGATCAAGGACTTTTATAATTCACTCTTTTGAGAACCAA 3813  
Db 4372 GGCCTTGGTGGAAATACAGGTTCAGATCCAGACCGGTGATTTACAAGAACTAGAGAACTCAG 4431  
QY 3814 GACAGCTGCCCCCAACAGGTTCAGAAACAGCCCTGTGTACAGGAGCAATGTTATCAC 3873  
Db 4432 CATCTGTGCCCAGAGCAGATGTTAGAAACAAAAATCATGTTATGATGGACAGTGTCTATGA 4491  
QY 3874 TACATATGAAAGCAAGTCTTTGGAACAAATAAGAACTGTATGTCGCCAGCTTCA 3933  
Db 4492 TATAAATGGAATGGCCAGTGTCTTGGAGGGCTTCCCGAACAGTGTGTGTCAAAAGTCA 4551  
QY 3934 GATGCGTTAATGTACAGAGGCTGTCTCCCTCAGCCCGCTCTGTCTGCTGCAATTCGGCAG 3993  
Db 4552 GATGATATAAATGTAAACAGGGGGCTGTCTTGGTGTATGAGCCAGCCCTGATGCCAGAGTCT 4611  
QY 3994 TGCAATCCAGCTTCAGAAACCTTCTCTCTACTGTACACAGGCTGGAGTCTGTGGTTGT 4053  
Db 4612 TGTAAACCACCGTGTGTGTAAACCCCACTGTGTGTGTAGCGAGACAAAACATGCCATTTGT 4671  
QY 4054 GAGAAGGCTTATACAGAGATAATGAAATCAAAATGTTTCTCTGGAATTAAGTCA----- 4105  
Db 4672 GAAGAAGGTTACATCTGAAGTCTATGCTTCTTAAACAGCACCTCTTGAGCAATGCACACTTATC 4731  
QY 4106 -----TGAAAGTACAGGCTCAGAGATTAATAAGCTGATGTGAAAAACCTTTCTG-G 4157  
Db 4732 CCCCTGTGTGTATTTACCCACCAATGGAGGACAAAAGAGGAGATGTGAAAAACAGTCGGCT 4791  
QY 4158 GAAAAACAGACCTGTGAATTTCAAAAAATACATGATATTTTAAAGGA--TGCTCTCTTCAA 4215  
Db 4792 GTACATCAACCCAAACCTCCAGTAAACCCAGGAGCGGGAGGAGACTGTGTTTCTACAG 4851  
QY 4216 CCACTTGATCCAGATGCGGAGTAAAAATTTGGGTTTATGGCGTTTCAGGTGGCGCTTTT 4275  
Db 4852 CCAATTTGGCCAGATGGAGACTAAAGACCTGGGTTTACGGTGTAGCAGCTGGGGCATTT 4911  
QY 4276 CTCATCATGATTTTCTTAATATTTACTTCTCTACCTTGTTCGAGAGAGCCA---AAACCA 4332

Db 4312 GTGTTACTCATCTTTATTTGTTCTCCATGATTTATCTAGCTTGCAAAAAGCCAAAGAACCC 4371  
Qy 4333 CATCAAGACACACCTCCCAACAGAGCCTCTGACCTTAGCTACGATGAGACTTAGAC 4392  
Db 4972 CAAAGAGGCAAAACACCGACTGAACCTTTAACTTTAGCTATGATGAGATGCCGAC 5031  
Qy 4393 ATGTAA 4398  
Db 5032 ATGTAA 5037

RESULT 8  
AX818221  
LOCUS AX818221 5032 bp DNA linear PAT 10-DEC-2003  
DEFINITION Sequence 92 from Patent WO03068268.  
ACCESSION AX818221  
VERSION AX818221.1 GI:39723273  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1  
REFERENCE  
AUTHORS Ek,S., Borrebaeck,C.A. and Ehinger,M.  
TITLE Treatment&comma; diagnosis and imaging of disease  
JOURNAL Patent: WO 03068268-A 92 21-AUG-2003;  
BIOINVENT INTERNATIONAL AB (SE)  
FEATURES  
source location/Qualifiers  
1..5032  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

ORIGIN  
Query Match 25.7%; Score 1129.6; DB 6; Length 5032;  
Best Local Similarity 58.5%; Pred. No. 0;  
Matches 2265; Conservative 0; Mismatches 1524; Indels 81; Gaps 14;  
Qy 610 TGGAGCCCTCTCTCAAGACATGCGCTTTCAGGAGTCTCTGCGAGATTTAGGAGCG 669  
Db 4 TGGAGCCCTCTCAAAACATGCCATGATGTCCTCCCTCGAGCATCTGTTAAGG 63  
Qy 670 AGCCGGAACGTGAACACATGGCTATTGGAGGTGGAAAGAGTGTCTGAACCTTTGAG 729  
Db 64 ACACGAACCATCAGGAGTTCCTCAATGGCAGTGAAGAGGAGTTCAGAAATTTGAAG 123  
Qy 730 AAAGAGGCTGCTATGTTGAAGGAGAACTTCTGCGAGCAATGTCAGGTATTCCTGAGA 789  
Db 124 AAAGAACCTGTTGTTCTCAGGAGATGGAGTTCCTCCCTGTCAGCTATGCTGAGA 183  
Qy 790 ACTTCTGAATGAAGAAATGCAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 849  
Db 184 ACTACAGAGTGGACTGAGTGCCTGTGGAGCCCTTTGCTCAGTCAGCAGGACAGAGGCGC 243  
Qy 850 CATGTGAGGAGCCGCTGTGGCGGTGGATCCAGACCGGAGGTGTACTGTGCCAG 909  
Db 244 GCGAACAGACGCGCTCTGTGGAGGGGGCATCCAGACCCGAGAGGTGTACTGCGTGCAG 303  
Qy 910 -----AGCGTACCAGCAGCTGCGCGCACTGAGGGCCAAAGAGTCTCTACACCT 957  
Db 304 GCCAACGAACACCTCTCTCAATTAAGTACCCACAGAACAAAGAGCTCAAGGCCA 363  
Qy 958 GTGAAAGGCAATATGTGTGGAGCCCGCGCTGTGCCCTCTCAGCTCTGCAATATCCCT 1017  
Db 364 ATGGACTTAAATATATGCACTGGACCTATCCCTTAATACTACACAGTGTGCCACATTCCT 423  
Qy 1018 TGCTCTACGAGTGCATAGTATCTCTCTGTCAGCTGGGGCTGTGTCATCCATGAAGAAC 1077  
Db 424 TGTCCTACTGATGTGAGTTTACCTTGTGTGCTGAGCTTGGGACCTGTACTTATGAAG 483  
Qy 1078 TGTCATGAACCTCAGGGGAAAAAGGATTTAGAACGAGGAGCG-----CCATGTC 1128  
Db 484 TGTAAATGATCAGCAAGGAAAAAGGCTTCAAACTGAGGAAGGGGGGCAATTACCAATGAG 543

Qy 1129 CTCTAGGAATCTACAGGCGCTGCAGGCAATTCCTCAATTTGGTGGAGTCTGTCTCTGT 1188  
Db 544 CCCACTGGAGGCTCTGGGGTAACCGAACTCTGCTCACTTACTTGAAGCCATTCCTGT 603  
Qy 1189 GAGGATCCAAATGTGTCTACGATGGCTGGCAT---CAGAAAGGATCTGTTTCCCTGATCAT 1245  
Db 604 GAAGAGCCCTGCTGTTATGACTGGAAAGCGGTGAGACTGGGAGACTGCGAGCCAGATAAC 663  
Qy 1246 GGAA---AATGTGGCTGGGACATCGTATTCTGAAGGCCGTCTGCCAGATGACCGCGGA 1302  
Db 664 GGAAGGAGTGTGGTCCAGGCCACGCAAGTTCAAGAGGTTGTGTGCATCAACAGTATGGA 723  
Qy 1303 GAAGATGATCAGGAGTCTTTGCCAGTTCCCTCTCTCTCTGAGAGGAGTCTTGTGAA 1362  
Db 724 GAAGAAGTTGACAGACAGCTGTGCAGAGATGCATCTTCCCATCCCTGTGGCTGTGAT 783  
Qy 1363 ATTCCCTGCCGAATGCACTGTGTGCTGAGCGAGTGGACGGAGTGTCTATCTCTTCCAG 1422  
Db 784 GCCCATGCCCCGAAGACTGTGTCTCAGACATGCTTACGTGTCTCTCTGTCTCACAC 843  
Qy 1423 TCCTGTTCAAATAAAACTCAGATGGGAAACAGACACAGTCAAGAACTATCTCTGC---A 1479  
Db 844 ACCTGCTCAGGGAAACGACAGAGGAAACAGATACGAGCAGCATTCATTCTGSCCTAT 903  
Qy 1480 CTGGCTGGGAGGTGGAAAGCCATGTCCTCTAGTCTCCTCAGGCTCTCAGAGCATCTGTTG 1539  
Db 904 GCGGGTGAAGAGGTGGAAATTCGCTGTCCAATAGCAGTCTTTTCAAGAAAGTACGAAGC 963  
Qy 1540 TGTAAATGACCATTCCTGTATGACGCTTCACTGGGAGACATCGCTTGGGGCCCTTGTCT 1599  
Db 964 TGTAAATGACCATTCCTGTACAGATGTACCACTGGCAACTGTGTCTCTGGGCGCAGTCAT 1023  
Qy 1600 GAGGACATTTGGTAACTGCTTAACTGCAACCACTGGCTGGAATGGAGAGCAACGCTGT 1659  
Db 1024 GAGGACACTCAGTATCTCTTCAACAACACTAGCTTGGAAATGGGAGGCTCTCTGC 1083  
Qy 1660 GGTGTAGGCATTACAGCTCGGAGAGTCTTGTGTCAAGAGTCACTGGGCAACAGTAATG 1719  
Db 1084 TCTGTGCGCATGCAGCAAGAAAGTCACTGTGTGCGAGTCAATGTGGGCAAGTGGGA 1143  
Qy 1720 ACCAAAGATGTCAGATTTACTGCACTGAAACTGTGCGCCCTCTGTTTCTCTCCATGC 1779  
Db 1144 CCAAAAAAATGTCTGAAAGCCTTCGACCTGAAACTGTGAGGCTTGTCTCTCTCTGT 1203  
Qy 1780 AAAAAGACTGTATTGTCACTGCTTTCAGTCAGTGGGACCCCTGCCAAGAGATGCGCAA 1839  
Db 1204 AAGAAGACTGTATTGTGACCCCATATAGTACTGGACATCATGCCCTCTCTCTGTGAAA 1263  
Qy 1840 GCAGAAATGCCACAGT---AAAAAGTCTCTGATACAGAAATCATCTCCAAAGAGCAGCC 1896  
Db 1264 GAAGGGGACTCCAGTATCAGGAAGCAGTCTAGGCATCGGGTCACTTCAGCTGCCAGCC 1323  
Qy 1897 AATGAGGCCAGGAATGCCAGATACCTTATATAGGAGAGAGTGTGAGATGTTTCC 1956  
Db 1324 AACGGGGCCGAGACTGCACAGATCCCTCTTATGAAGAGAGGCTGTGAGGCACTCCAA 1383  
Qy 1957 TTGTCTCTGTATTCGGTGGAAAGCCACAGAAATGGAGCCCTTGCATCTTAGTGCCAGAG 2016  
Db 1384 GCGTGCCAAAGCTACAGTGGAGAGCTCACAATGGCGCAGATGCCAATTAGTCTCTTGG 1443  
Qy 2017 TCTGTCTGGAGGAAATGAACCGGCGAGCTGTAAGCTGTGGAAGGGGTTCACAAAGAA 2076  
Db 1444 AGCGTGCAACAGACAGCCCTTGGAGCAGAGAGGCTGTGGGCTGGGCGCAGAGCAAGA 1503  
Qy 2077 GCTGTCTCATGCTCTCTGATGACAAACCGGTGACAGAAATGATGGAATGCTCAAGCAG 2136  
Db 1504 GCCATTACTTGTGCGAAGCAAGATGGAGCAGGCTGGATCCATGATGTGCTTACAGTAT 1563  
Qy 2137 ACAAAGGCATGCTCTCTCTTGTGCAAGAAATGCAAGTCCCATCTCGAGAGAGTGCACC 2196  
Db 1564 GCGGCCCTGTGCCAGCCCTTACCAGGCCCTGCCAGATCCCTGCCAGGATGACTGTCAA 1623



Db 3784 GCTTCAAAAGCCAAAGAAAGCCCAAGAGGCAAAACACGACGACCTGAACCTTTAACC 3843  
QY 4369 TTAGCCTACGATGGAGACTTAGACATGTAA 4398  
Db 3844 TTAGCCTATGATGGAGATGCCGACATGTAA 3873  
RESULT 9  
AX780247  
LOCUS AX780247 5287 bp DNA linear PAT 14-JUL-2003  
DEFINITION Sequence 2404 from Patent WO03039443.  
ACCESSION AX780247  
VERSION AX780247.1 GI:32697241  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
REFERENCE  
AUTHORS Haferlach, T.; Schoch, C.; Kern, W., Kohlmann, A., Schnittger, S.,  
Dugas, M., Eils, R., Brors, B. and Mergenthaler, S.  
TITLE Novel genetic markers for leukemias  
JOURNAL Patent: WO 03039443-A 2404 15-MAY-2003;  
Deutsches Krebsforschungszentrum (DE);  
Ludwig-Maximilian-Universitaet Muenchen (DE); Haferlach, Torsten,  
PD Dr. Dr. (DE); Schoch, Claudia (DE); Kern, Wolfgang (DE)  
FEATURES  
source  
1..5287  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Query Match 25.2%; Score 1109.8; DB 6; Length 5287;  
Best Local Similarity 58.4%; Pred. No. 0;  
Matches 2260; Conservative 0; Mismatches 1529; Indels 83; Gaps 16;  
QY 610 TGGAGCCCTGCTCCAGACATGGCTATTGGAGGTGGAAAGGAGTGTCTGAACTCTTTGAG 729  
Db 4 TGGAGCCCTGCTCAAAACATGCCATGACATGTGTGCCCTGAGGACCTCGTGTAAAG 63  
QY 670 AGCGGAACGTGAAGACATGGCTATTGGAGGTGGAAAGGAGTGTCTGAACTCTTTGAG 729  
Db 64 ACAGAACCATCAGCAGTTCCTCCATTTGGCAGTGAAGAGTGTCCAGATTTGAAGAA 123  
QY 730 AAAGAGGCTGATTTGAAAGGAGAACTTCCTGAGCAAGTCCCGAGTATCTCTGGAGA 789  
Db 124 AAAGAACCTGTTTGTCTCAAGGAGATGGAGTGTCTCCCTGTGCCAGCTATGGCTGAGA 183  
QY 790 ACTTCTGAATGAAAGATGCCAGTCTCTCTCCTCGAGCAGCAGGATCCCCACTGG 849  
Db 184 ACTACAGAGTGAGCTGAGTGCCTGTGAGACCTTTGCTCAGTCAGCAGGACAAGAGCGC 243  
QY 850 CATGTGACGGACCCGCTGTGTGGGGTGGGATCCAGACCCGGGAGGTGTACTGTGCCAG 909  
Db 244 GCGAACCAAGCGCCCTCTGTGGAGGGGATCCAGACCCGAGAGGTGTACTGTGTCAG 303  
QY 910 -----ACGCTACAGCAGCTCCCGACTGAGGCGCAAGAGTCTTAGACCT 957  
Db 304 GCCAACGAAACCTCCTCTCAATTAAGTACCACAGAAACAAGAGGCTCAAGCCA 363  
QY 958 GTGGAAGAGGCATTATGTGTGGACCCGCCCTGTGGCCCTCTCAGCTCTGCAATATCCT 1017  
Db 364 ATGGACTTTAAATATTGACATGGACCTATCCCTAATACTACAGCTGTGCCATTTCT 423  
QY 1018 TGCTTAGGACTGCATAGTATCTTCCTGTGTCAGCCTGGGCGCTGTGCATCCATGAAAC 1077  
Db 424 TGTCCAACGTGAATGTGAAGTTTCACTTTGGTTCAGCTTGGGACCTTTGCTATATGAAAC 483  
QY 1078 TGTATGAACCTCAGGCGAAAGAAAGATTTAGAACGAGGAGCG-----CCATGTC 1128  
Db 484 TGTATATGATCAGAGGAAAGAAAGGCTTCAACTGAGGAAGGCGGATACCAATGAG 543

QY 1129 CTCATGGAATCTACAGGGCTCGAGGCAATGCCCTCATTTGGTGGAGTCTGTTCTTGT 1188  
Db 544 CCACCTGGAGGCTCTGGGTAAACCGAAACTCCCTCTCACTTACTGGAAGCCATTCCTTGT 603  
QY 1189 GAGGATCCAATGTCTACCGATGGCTGGCAT---CAGAAGGATCTGTTTCCCTGATCAT 1245  
Db 604 GAAGAGCCTGCCTGTTATGACTGGAAAGCGGTGAGACTGGGAGACTGCGAGCCAGATAAC 663  
QY 1246 GGAA---AATGTGGCTGGGACATCGTATTCTGAAGGCGGTCTGCCAGATGACCGCGGA 1302  
Db 664 GGAAAGGAGTGTGGTCCAGGCGACCAAGTTCGAAGAGTGTGTGTCATCAACAGTGAATGA 723  
QY 1303 GAAGATGTATCAGGGAGTCTTTGCCCCAGTTCCTCCCTCTCTGAGAGGAAGTCTTTGTAA 1362  
Db 724 GAAGAAGTTGACAGACAGCTGTGCAGAGATGCCATCTTCCCATCCCTGTGSCCTGTAT 783  
QY 1363 ATTCCCTGCCGAATGAGCTGTGTGTGAGCGAGTGGAGCGAGTGTCTATCTCTGTTCCAG 1422  
Db 784 GCCCATGCCCGAAAGACTGTGTGCTCAGCACATGCTGTACGTGTCTCTCTGCTCAGAC 843  
QY 1423 TCTGTTCATAATAAACTCAGATGGGAAACAGACAGCTCAAGAACTATCTCTGCG---A 1479  
Db 844 ACTGTCTAGGGAAGAAACGACAGAGAGGAAACAGATACGAGCAGATCCATTTGSCCTAT 903  
QY 1480 CTGGCTGGGAAGGTGGAAGCCATGTCTCCCTCTAGTCAAGTCTCTCAAGAGCATCGTTG 1539  
Db 904 GCGGGTGAAGAGGTGGAATTCGCTGTCCAAATAGCAGTGTCTTGCAGAGAGTACGAAGC 963  
QY 1540 TGTATGACCATTCCTGTATGAGCTTCTCAGCTTGGGAGACATCGCTTGGGCGCTTGTCT 1599  
Db 964 TGTATGAGCATCCTTTGCACAGTGTACCATCTGCAAACTGCTTTCGAAAGAGTACGAAGC 1023  
QY 1600 GAGGACATATGTTAACTGCCCTTAATGCAACCATTTGGCTGGATGGAGAGCAACGCTGT 1659  
Db 1024 GAGGACACCTCAGTATCGTCTTCAACACACTACGACTTGAATGGGAGGCTCTCTGC 1083  
QY 1660 GGTGTAGGCAATTCAGACTCGGAGAGTCTTCTGTGTCAAGAGTCACTGTGGAAGTATG 1719  
Db 1084 TCTGTGGGATCGACAGCAAGAAAGTCTATGTGTGCGAGTCAATGTGGGCGCAAGTNGA 1143  
QY 1720 ACCAAAGATGTCCAGATTTCTACTCGACTGAACTGTGGGCGCTCTTCTTCCCATGC 1779  
Db 1144 CCCAAAGATGTCTTGAAGCCCTTCGACTGAACTGTGAAGCCCTTGTCTGTTCTTGT 1203  
QY 1780 AAAAAGACTGTATTGTGACTGCTTTTCACTGAGTGGACACACCTGCGCCCAAGAT-GTGCA 1838  
Db 1204 AAGAAGGACTGTATTGNNNNCCATATAGTGACTGGAACATCATGCCCTCTTCTGTGTA 1263  
QY 1839 AGCAGGAATGCCACAGT---AAAACAGTCTCGATACAGATCATATCCCAAGAGCAGC 1895  
Db 1264 AGAAGGGACTCCAGTATCAGGAAGAGTGTAGGATCGGGTCAATCTCAGCTGCCAGC 1323  
QY 1896 CAATGAGGCCAGGAATGCCAGATACCTTTATATGAGGAGAGAGTGTGAAGATGTTTC 1955  
Db 1324 CAAACGGGCGCGAGACTGCACAGATCCCTCTTATGAAGAGAGGCTGTGAGGCACTCA 1383  
QY 1956 CTTGTGCTCTGTATATCGTGGAGGCCACAGAAATGGAGCCCTTGCATCTTAGTCCAGA 2015  
Db 1384 AGCGTGCCAAAGCTACAGTGGAAAGTCACAAAATGGCGCAGATGCCAATTAGTCCCTTG 1443  
QY 2016 GTCTGTCTGGCAGGAATAAACGGGCGAGCAG-TGAAGCCTGTGGAAAGGGGTACAAACAA 2074  
Db 1444 GAGCGTGCACAAAGACAGCCCTGAGCAGCAGGAGGCTGTGGGCTGGGCGCAGGCA 1503  
QY 2075 GAGCTGTCTCATGCAATCTCTGATGACACCGGTGAGCAGAAATGATGAAATGCTTCAAGC 2134  
Db 1504 GAGCCATTAATTTGTGCAAGCAAGATGGAGCAGGCTGGAATCCATGAGTGCCTACAGT 1563  
QY 2135 AGACAAAAGGCATGCTCTCTTGTGCAAGAAATGACAGTCCCATGTCGAGAAGACTCA 2194  
Db 1564 ATGAGGCCCTGTGCGAGCCCTTACCGAGGCTTCAGATCCCTTCCAGGATGACTGTC 1623  
QY 2195 CCTTCACTGCTTGGTCCAAAGTTTACGCCCTTGTCTCCAGAACTGTGAAGCCCAAAAGTA 2254



Db 1624 AATTGACCAAGTGTGCAAGTTTCTTCATGCAATGGAGACTGTGGTGCAGTTAGGACCA 1683  
QY 2255 GCGCGCAGAGCTCACAGGAAAAGCAGAAGAGGAGAAATCCAGGATTTCTGACCTTT 2314  
Db 1684 GAAAGCCGACTCTTGTGTGAAAAGTAAAGAGGAGGAAAATGTAAAAATCCCACTTGT 1743  
QY 2315 ACCCTCTAGTGGAGACAGAACTATGCTCTGTGATGAATTTATATCCCACTTATGGAA 2374  
Db 1744 ATCCCTGATGAGACTCAGTATTTGCTTGTGACAAATATAATGACAACTGTGGGA 1803  
QY 2375 ACTGGTCAGATTCGATCTTCCAGAGGAGAGGAGCCCTCACGAGGACTCGGGTAC 2434  
Db 1804 ACTGGTCAGACTGTATTTACCAGAGGAAAAGTGGAAAGTGTCTGGGAATGAAAGTAC 1863  
QY 2435 AAGCAGACACCAAGATGTGGAGAGGCGCTGCGCTTTCAGCAGTAGCTGTCTGATA 2494  
Db 1864 AAGAGACATCAGGAATGCGGACAGGATATGTTTACCAGCAATGGCATGCTACGATC 1923  
QY 2495 AAAATGGAAGACCTGTTGACCCCTCTCTGCGACAGCTCTGGTTACATTTCAAGAAAAAT 2554  
Db 1924 AAAATGSCAGGCTTGTGAAAACATCTAGATGTAAACAGCCATGTTTACATTTGAGGAGCCT 1983  
QY 2555 GTGTCACTCCCTGCCCATTTGATGCGAATTAAGCAATGGTCTAGTTGGGGTCTTGCA 2614  
Db 1984 GCATCATCCCTCGCCCTCAGACTGCAAGCTCAGTCAGTGTGCTCCAACTGTGCGCTGCA 2043  
QY 2615 GTTCATCTTGTGAAATGGAGTGGAAATTCGATCCAAATGCGTAAAGAAAACCTTACA 2674  
Db 2044 GCAAGTCTGTGGAGTGTGTGAAGTTCGTTCTAAATGCGCTGCNGAANWCCATATA 2103  
QY 2675 ATGAGGACGACCAATGTCCTCCAACTGGATC-----TCAGAAATCAGGTACATGAGGCAG 2728  
Db 2104 ATGAGGAGAGGCTTGGCCCAACTGGACCATGTCAACACGAGGCACAGGTGTATGAGGTTG 2163  
QY 2729 TCCCATGTTACAGTGAATCAATCAGTATTTCTCGGGTGTAGAACCTGCTTCTATGCA 2788  
Db 2164 TCCCATGCCACAGTACTGCAACAGTACTATGCGNCAAGAGCCCTGGAGCATCTGCA 2223  
QY 2789 AAATCAACAAATGAGCTGAGTCCCTCGCTGTGGAGGAGGAAACAATCTTAGGAAATCA 2848  
Db 2224 AGGTGACCTTTGTGAATATCGGGAGAACTGTGGAGAGGCGTGCAACCCGAAAGTGA 2283  
QY 2849 GATGTGT---GAATCTGCGGATGGTGAAGGTGGAGAGTGGATAGCAACCTGTGCAAC 2905  
Db 2284 GATGCAATGCAAGAAATACAGAGATGGCCCTTCTTAAACATGTAGAGGATTAACCTTNTNACC 2343  
QY 2906 AGGATGAAATTTCCCCAGAAACCCAGTCTGTCTTATGCTGCCAATGAGTGTGCA 2965  
Db 2344 CAGAAGAGATGCCCTGGGCTCTAGAGTGTGCAAAATACCATGCCCTGAGGACTGNTGA 2403  
QY 2966 TGTCTAGTGGGACTTTGGAGCAAAATGCCCAAGTCATGCCGATCCCCACAAAT---GC 3022  
Db 2404 TATCTGANTGGGTCCTATGAGACCAATGNTTTTGGCTTGCATCAAGCAGTTTCCGCGC 2463  
QY 3023 AGAGAGAACTCGCCACCTGCTAAGACCTACATGACTCAAGGACTTGTGCTGAAGACT 3082  
Db 2464 AAAGTCAAGTATGCCATCAGACACAGAGCTGATGAAGGAAGATCTTGCCCTAATGCTG 2523  
QY 3083 CACAGGTGACGCTTGCTCTGAAATGAAAATGCTTCCAGTTCCAGTACAATCTAAACAG 3142  
Db 2524 TTGAGAAAGAACCTGTAACTTGAACAAAACCTGCTACCACTATGATTAATATGTAACAG 2583  
QY 3143 AGTGGAGCACAATCCAGCTGAGTGAAGAACGACCCCTGTGGTCAAGGCGTCAGGACCCGC 3202  
Db 2584 ACTGGAGTACATGCTAGCTGAGTGAAGGAGCAATTTGTGGAATGGAATPAAAAACAAGGA 2643  
QY 3203 TGTAAAGCTGTGTGTGAGTGAAGGACAGCTCAGCATGAGCAATGTCAGGAGTGAAGCAGATA 3262  
Db 2644 TGTGTGATTTGTTCGAAGTGAAGGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 2703  
QY 3263 ATTTGAGAGAGCCCAAGAAATGAGCAATTCCTCGCTTGTGGAAATGCGTGTCAACTGTC 3322

Db 2704 GCTTGGAGAGAACTGGCAGATGAACAGCTCCTGATGGTGGAAATGCCCTGTGAATGTC 2763  
QY 3323 AGCTCTCAGGCTGGACGCTTTGGACAGAGTGTTCACAGACCTGTGGCCATGGAGGTGAA 3382  
Db 2764 AGCTTCTGATTTGGTCTCCTTTGGTCAAGTGTCTCTCAAACATGTGGCTCACAGAAAA 2823  
QY 3383 TGAGCCGACCTCGATTTATCATTTATGCCAAACCCAGGAGAGGACGCGCCATGCCACAG 3442  
Db 2824 TGATCCGAGACGAAACAGTGAACCCAGCCCTTTCAAGGTGATGGAAGACCATGSCCTTCCC 2883  
QY 3443 AGCTTACCACGAGAGAAAACTCTGCCAGTGAACCCCTCTGTACAGCTGGGTCTTGGCAACT 3502  
Db 2884 TGATGGACAGTCCAAACCCCTGCCAGTGAAGCTTGTATCGGTGGCAATATGGCCAGT 2943  
QY 3503 GGTCTGCACTGAATTTGGAGGCTGGAAGTGTGGGAAAGGATTCAGATCCGAGCCCTTT 3562  
Db 2944 GGTCTCCATGCCAATGTGAGGAGCCCGAGTGTGAGAGGAGGACCAGAACAAAGAACATTT 3003  
QY 3563 CCTGCATGCTCCACAGTGGTTCAATATCTCATGAGCTGGACGTGTGAGATGCACGTG 3622  
Db 3004 CTGTGTAGTAAAGTGTGGTCACTGATGATTTTCAGCAAGTGTGTGATGAGGAATCT 3063  
QY 3623 GTGGAGAAATGCCCTT-----TCAGGACAGCATCCTGAAGCAGCTGTGTT 3667  
Db 3064 GTGCTGACATTTGAATCATTTATAGATGGTAAATAAAATATGGTTCTGGAGGAATCTTGCA 3123  
QY 3668 CTGTGCCCTTGGCCAGGAGACTGCCATTTACAGAAATGGTTCAGAGTGGAGCACATGTGAT 3727  
Db 3124 GCCAGCTTGGCCAGGTGACTGTTATTTGAAGGACTGTCTTCTCTGGAGCCTGTGTGAGC 3183  
QY 3728 TAACTGCAATTTGATGGAAGAGCTTTGAGACTGTGGCCGCGCAGTCTAGATCAAGGACTT 3787  
Db 3184 TGACTGTGTGAATGGTGAAGTCTAGGCTTTGGTGGAAATACAGGTTCAGATCCAGACCGG 3243  
QY 3788 TTATAAATTCAGTCTTTTGGAGAACCAAGACAGCTGCCCCCAACAGGTTCTAGAAAACAGGCC 3847  
Db 3244 TGATTTACAAGAACTAGAGAAATCAGCATCTGTGCCCAGAGCAGATGTTAGAAAACAAAT 3303  
QY 3848 CTTGTACAGGAGGCAAAATGTTATCACTACATCGAAAGCAAGTCTTTGGAAACAATAACG 3907  
Db 3304 CATGTTATGATGGACAGTGTCTATGAATATAAATGATGGCCAGTCTTTGGAGGGCTCTT 3363  
QY 3908 AACGAACTGTATGTTGCCAGGCTTCAGATGCGGTAAATGTCAAGGAGGTGTCTCCCTC 3967  
Db 3364 CCCGAACAGTGTGGTGTCAAAAGGTCAAGATGATAAAATGTAAACAGGGGGGTGTGTTGTA 3423  
QY 3968 AGGCCGCTCTGCTGCCATTCGGCAGTGCATTCAGCCTTCAGAAAPACCTTTCTCTCTACT 4027  
Db 3424 TGAGGACCTGATGCGACAGGCTTGTAAACCCCGTGTAGTCAACCCCACTCTGTA 3483  
QY 4028 GTACACAGGCTGAGTCTGTGTTGTGAGAGGCTATACAGAGATAATGAAATCAAAATG 4087  
Db 3484 GTAGCGAGACAAAAACATGCCATTTGTGAAGAGGGTACACTGAAGTCAATGCTCTTCTAA 3543  
QY 4088 GTTTCCTGGATTACTGCA-----TGAAAGTACAGGCTCAGAGGTAAAG 4132  
Db 3544 GCACCTTGAACAATGACACACTTATCCCGTGTGGTATTACCCACCATGGAGGACAAA 3603  
QY 4133 AAGCTGATGTGAAAAACCTTTCTG-GGAAAAACAGACCTGTGAATTCAAAAATACATGAT 4191  
Db 3604 GAGGAGATGTGAAAAACAGTCCGCTGTACATCAACCCCAACCTCCAGTAAACCCAGCAG 3663  
QY 4192 ATTTTAAAGGA---TGTCTCTTCAACCACTTGAATCCAGATGGCCGAGTAAATTTGGG 4249  
Db 3664 GACGGGAAGGACCTGTTCTTACAGCCATTTGGGCCAGATGGGAGACTAAAGACCTGGG 3723  
QY 4250 TTTATGGCTTTCAGGTGGCGCTTTCTCATCATGATTTTCTTAATTTTCTTCTTACC 4309  
Db 3724 TTTAGGTGTAGAGCTTGGGCACTTGTGTACTCACTTTTATGTTCTCCATGATTTATC 3783  
QY 4310 TTGTTTCAAGAGGCCA---AAACACATCAAGACACCTCCCCCAACAGAGCCTTGA 4366  
Db 3784 TAGCTTCAAAAAGCCCAAGAAACCCCAAGAGGCAAAACCAACCGACTGAAACCTTTAA 3843

QY 4367 CCTTAGCTACGATGGAGACTTAGACATGTAA 4398  
|||||  
Db 3844 CTTAGCTATGATGGAGATCGGACATGTAA 3875

## RESULT 10

AX747433

LOCUS AX747433 3053 bp mRNA linear PAT 20-JUN-2003  
DEFINITION Sequence 958 from Patent EPI308459.

ACCESSION AX747433

VERSION AX747433.1 GI:32131821

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## REFERENCE

1 Isogai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S.,  
Yamamoto, J. I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R.,  
Tamechika, I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahari, K. and  
Masuho, Y.

Full-length cDNA sequences

Patent: EP 1308459-A 958 07-MAY-2003;

Helix Research Institute (JP); Research Association for

Biotechnology (JP)

Location/Qualifiers

1. 3053

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

## FEATURES

source

Query Match 13.2%; Score 580.6; DB 6; Length 3053;

Best Local Similarity 56.4%; Pred. No. 3.9e-165;

Matches 1309; Conservative 0; Mismatches 934; Indels 78; Gaps 9;

## ORIGIN

QY 10 TGCATTGAGAGCTGAAACCGAACTGTGGTTGCAATGAAATATGCGAACACTTTGCCCTT 69  
Db 730 TGCAATCCAGAAAGACAAAGACATTCCTCGGAGGATATCATCTGTAGTACTTTTGAGCCC 789  
QY 70 CAGCTCTCTACAGACAGCTTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 129  
Db 790 AAGCT 849  
QY 130 TCTTACCATGTCACATGCTAGCAAGGATGTGGGAAGAAATTCAGCATAGAACTCGC 189  
Db 850 TTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 909  
QY 190 GCGGTACATGCTCCCT 249  
Db 910 CATGTGGTGGCGCTCCCGAGTTCGAGGCTCTGCTGCTCCAAACCTTACGAGGTTCCAG 969  
QY 250 GCGTGTGATGCTCCCATTTCT 309  
Db 970 GTGTG-----CCAAATCCAGTCCATCGGAGGCGGAGGAGCTCAGGTACAGCTGCAT 1020  
QY 310 GTTGAACCATGAGTAAATGCACTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 369  
Db 1021 GTGGGCGCTTGGAGCACCTGCTCAATGCTCCCACTCCGACCAAGTAAGCAAGGAGGA 1080  
QY 370 ACTGTTCTGATTTTAA-----CTCTGATTCAAATGAG 402  
Db 1081 CCGCGGAGATAAAGAACGGGAAGAGACCGCAGCAAGAGTAAGGATCCAGAGCC 1140  
QY 403 CGAGTCACCTTTAAACATCAAGTTTACA-----AAGCACATCATTCATTCGAAGTCT 453  
Db 1141 CCGCAGCTTTATTAAGAAAGAAAGAAACAGGACAGGACAGCAAGAGAGAAATAT 1200  
QY 454 TGGGCAATAGACATAGGTTTATCAACCCGCGAGGTTTCGTGTACAGAGAGTGTGACAA 513  
Db 1201 TGGGACATCCAGATTGGATATCAGCCAGAGGTTTATGCAATTAACAGACCGGGAAA 1260

QY 514 AATGCTATGTTAAGCTTTTGCCCTTCAAGATTCCTTCCCATGACTGTTTCACTGCTGCATC 573  
Db 1261 GCTGCTGATTTAAGCTTTTGGCAGCAAGAGAGACTTCCAATGACCTTCCAGCTCTGTGTG 1320  
QY 574 ATGCCCAAGAGCTGTGAACCTCCAGTGTCTCTCTCTGAGCCCTGTCTCAAGACATGC 633  
Db 1321 ATCACCAGAGAGTGCAGGTTTCCGAGTGTTCAGATGGAGCCCTGCTCAAAACATGC 1380  
QY 634 CGTTCAGGAGTCTCTTTCAGGATTTAGGACAGAGCCGGAACGTGAAGCACATGGCT 693  
Db 1381 CATGACATGTTCTCCCTCGCAGGCTCTGTTGAAGACACGAACCATCAGGAGTTTCCC 1440  
QY 694 ATTGAGAGTGAAGAGAGTCTCTCAACTTCTTGAGAAAGAGGCTCTGCTTGAAGA 753  
Db 1441 ATTGGCAGTGAAGAGAGTCTCAGATTTGAAGAAAGAACCTCTTTGTCTCAAGA 1500  
QY 754 GAATTTCTCAGCAATGTCCAGGATTTCTCGAGAACTTCTGAATGGAAGAAATGCCAA 813  
Db 1501 GATGAGATTTGCTCCCTGTGCCAGTATGCTGAGAACTTACAGAGTGGACTGAGTGCCT 1560  
QY 814 GTCTCTCTCTCTCTGAGCAGCAGGATCCCCACTGGCATGTGACGGACCCGCTGTGTGC 873  
Db 1561 GTGGACCTTTTCTAGTCAGCAGCAGCAAGAGCGCGGCAACACAGAGCCCTCTGTGA 1620  
QY 874 GGTGGATCCAGACCCGAGGAGTGTACTGTGCCAG-----AGCGTACAGCA 921  
Db 1621 GGGGGCATCCAGACCCGAGGAGTGTACTGCTGTCAGGCCCAACGAAACCTCTCTCACAA 1680  
QY 922 GCTGCCGCTAGGCGCAAGGAGTCTTACACCTGTGGAAGAGCATATGTGTGGA 981  
Db 1681 TTAAGTACCCCAAGAAACAAAGAGGCTCAAGAGCAATGGACTTAAATTTATGCACTGA 1740  
QY 982 CCGCGCCCTTGCCTCTCAGCTCTGCAATATCCCTTGTCTACGAGCTGCATAGTATCT 1041  
Db 1741 CCTATCCCTAATACACAGCTGTGCCACATCTCTTGTCCAACTGAATGTGAAGTTCA 1800  
QY 1042 TCTGTGTAGCTGGGCTGTGCTCATGAAACTGTCTATGAACTCTGAGGAGGAAAAA 1101  
Db 1801 CTTGTGTAGCTTGGGACCTTGTACTTATGAAACTGTGTAATGATCAGCAAGGAAAAA 1860  
QY 1102 GGAATTTAGAACAGGCGAGG-----CCATGCTCTCATGGAATCTACAGGCGCTGCA 1152  
Db 1861 GGTCTTCAACTGAGNAGCGCGCATACCAATGAGCCACTGGAGGCTCTGGGTGTTAC 1920  
QY 1153 GGCATTTGCCCTCATTTGGTGGAGTCTGTTCTTGTGAGGATCCAAATGTGTCACGATGG 1212  
Db 1921 GGAACCTGCGCTCACTTACTTGAAGCAATTCCTGTGAAGAGCTGCTGTATGACTGG 1980  
QY 1213 CTGGCA---TCAGAGGAGTCTGTTTCCCTGATCATGGAA---AATGTGGCTGGGACAT 1266  
Db 1981 AAGCAGTGAAGTGGGAACTGCGAGCCAGATAACGGAAGAGTGTGGTCCAGGACG 2040  
QY 1267 CGTATTTCAAGGCGCTCTGCCAGAAATGACCGGAGAGAGATGTATCAGGGAGTCTTTTC 1326  
Db 2041 CAAATTTCAAGAGGTTGTGTGCATCAACAGTGTGAGAGAAAGTTGACAGACAGCTGTGC 2100  
QY 1327 CAGATTTCCCTCTCTCTGAGAGGAGTCTTGTGAAATTCCTGCGGATGAGACTGTGTG 1386  
Db 2101 AGAGATGCCATCTTCCCATCCCTGTGGCTGTGATGCCCTTCCGAAAGACTGTGTG 2160  
QY 1387 CTGAGCGAGTGAAGGAGTGTCTCTCTGCTCCAGTCTCTGTTCAAAATAAAACCTCAGAT 1446  
Db 2161 CTCAGACATGTTCTACGTGCT 2220  
QY 1447 GGGAAACAGACCGAGTCAAGAACTATCTGTGGCACTGGCTGGG---GAGAGTGGAAAGCCA 1503  
Db 2221 GGGAAACAGATACGAGCAGATCCATCTGTGGCTATGCGGGTGAAGAGAGTGGAAATTCG 2280  
QY 1504 TGTCCCTCTAGTCAGGCTCTCAAGAGCATCTTGTGTGTAATGACCATTCCTGTGTATGAG 1563  
Db 2281 TGTCCAAATAGCAGTGTCTTGCAGAGATGACAGCTGTATGAGCATCTCTTGTGACAGTG 2340  
QY 1564 CTTACTGGGAGACATCGCCTTTGGGGCCCTTGTCTGTGAGGACACATTTGTAACCTGCCCTT 1623



Db 2341 TACACATGGCAAACTGGTCCCTGGGGCCAGTCGATTGAGACACCTCAGTATCGTCTTC 2400  
 QY 1624 AATGCAACCACTGGCTGGAATGAGAGAGCCACCTGTGGTGTAGGCATTGAGACTCGGAGA 1683  
 Db 2401 AACACAACCTAGACTGGATGGGAGGCTCTGCTGTCTGCGCATGAGACAAGAAA 2460  
 2Y 1684 GTCTTCGTCTCAAGAGTCAGTGGGACAAGTAATGACCAAAAGATGTCCAGATTTACT 1743  
 Db 2461 GTCATCTGTGTGGAGTCAATGTGGGCAAGTGGGACCCCAAAAATGTCTGAAAGCCTT 2520  
 2Y 1744 CGACCTGAACACTGTGGCCCTCTTTTCTCCCATGCAAAAAGACTGTATGTGACTGCT 1803  
 Db 2521 CGACCTGAACACTGTAGGCTCTGTCTCTCTGTGTAGAGAGACTGTATGTGACCCCA 2580  
 2Y 1804 TTCAGTGGAGTGGACACCTCCCAAGAGATGTGCCAAGCAGGAAATCCACAGT---AAAA 1860  
 Db 2581 TATAGTGACTGGACATCATGCCCTCTTCTGTGTAAAGAGAGGGACTCCAGTATCAGGAAG 2640  
 2Y 1861 CAGTCTCGATACAGAAATCATCATCCAGAGAGCCCAATGAGGCCAGGATGCCAGAT 1920  
 Db 2641 CAGTCTAGGATCGGTCATCATTCAGTCCAGCCAAACGGGGCGAGACTGCACAGAT 2700  
 QY 1921 ACCTTATATGAGGAGAGAGTGTGAAGATGTTTCCTTGTGTCTGTATATCGTGGAAAG 1980  
 Db 2701 CCCCTCTATGAAGAGAGGCTGTGAGGCACCTCAAGCGTGCCAAAGCTACAGGTGGAAG 2760  
 QY 1981 CCACAGAAATGGAGCCCTTCATCTTAGTCCAGAGTCTGTCTGGCAGGAAATACGGGC 2040  
 Db 2761 ACTCACAAATGGGCGAGATCCCAATTAGTCTTGGAGCGGTGCAACAAGACAGCCCTGGA 2820  
 QY 2041 AGCAGTGAAGCTGTGGAAGGGGTTACAAACAAGAGCTGTCTCATGCTCTCTGATGAC 2100  
 Db 2821 GCACAGAAGGCTGTGGCCCTGGCGACAGCAGAGCCATTACTTGTGCCAAGCAAGAT 2880  
 2Y 2101 AACCGTTCAGAGAAATGATGAATGGAATGCTCAAGCAGACAAACGGCAGTCTCTCTGTG 2160  
 Db 2881 GGAGGACAGGCTGGAATPCCATGATGCTACAGTATGACAGGCTGTGCGAGCCCTTACC 2940  
 2Y 2161 CAAGATGACAGTCCCATGTCGAGAGAGCTGCACCTTCACTCTGTTGTTCCAGTTTACG 2220  
 Db 2941 CAGGCTGCCAGATCCCTGCCAGATGATGCTCAATTGACAGCTGGTCCAAAGTTTCT 3000  
 2221 CCCTGTCTCCAGCACTGTGAAGCCACAAAAGTAGGGCGG 2261  
 3001 TCATGCAATGAGAGTGTGGTGAGTTAGCACCAGAAAGCG 3041

RESULT 11  
 AK092252  
 LOCUS Homo sapiens cDNA FLJ34933 fis, clone NT2RP7005669. PRI 15-JUL-2002  
 DEFINITION  
 ACCESSION AK092252  
 VERSION AK092252.1 GI:21750796  
 KEYWORDS oligo capping; fis (full insert sequence).  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1  
 AUTHORS Ishibashi, T., Kanehori, K., Yosida, M., Watanabe, S., Ishida, S., Ono, Y., Hotta, T., Hiraoka, S., Murakawa, K., Takiguchi, S., Kusano, J., Watanabe, M., Fujimori, K., Tanai, H., Ishida, M., Yamashita, H., Chiba, Y., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Takahashi-Fujii, A., Oshina, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Negahari, K., Masuho, Y., Nagai, K. and Isogai, T.  
 NEDO human cDNA sequencing project  
 Unpublished  
 2 (bases 1 to 3053)  
 REFERENCE 2  
 AUTHORS Isogai, T. and Yamamoto, J.

TITLE JOURNAL  
 COMMENT  
 FEATURES source  
 1. 3053  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="NT2RP7005669"  
 /cell\_line="NT2"  
 /cell\_type="teratocarcinoma"  
 /clone\_lib="NT2RP7"  
 /note="cloning vector: pME18SFL3-mRNA from NT2 neuronal precursor cells after 5-weeks retinoic acid (RA) induction."  
 ORIGIN  
 Query Match 13.2%; Score 580.6; DB 9; Length 3053;  
 Best Local Similarity 56.4%; Pred. No. 3.9e-165;  
 Matches 1309; Conservative 0; Mismatches 934; Indels 78; Gaps 9;  
 QY 10 TGCATTCAGAAGCTCAACCGAAGCTGTGGTTCGAATGAAATATCGAAGACACTTTGCCCTT 69  
 Db 730 TGCATTCAGAAGCAAGACATCTCTGGAGAGATATCATCTGTAGTACTTTGAGCCC 789  
 QY 70 CAGCCTCTCAAGAACAGCTTGGCTCATTCCTTGTCCCGGAGTTGTAGTATCTGAG 129  
 Db 730 AAGCCTCTCTCGAGCAGCTTGGCTCATTCCTTGGCAGCAAGATTGCATCTGTCTGAA 849  
 QY 130 TTCTTACCATGTCCTCACTGTAGCAAGGATGTGGAGAAATTCGACATAGAACCTGC 189  
 Db 850 TTTCTGCTGCTGTCGAATGCTCCAGACCTGCGGAGCGGGCTCCAGACCGGAGCGGT 909  
 QY 190 GCGGTCACTAGCTCCCTCTCTTTGGTGGTTGCAATGTCCAAATCTGACTGAGTCAAGA 249  
 Db 910 CATGTGGTGGCGCTCCGAGTTCGGAGGCTCTGGCTGTCCAAACCTGACGAGTTCAG 969  
 QY 250 GCCTGTGATGCTCCCATTTCTCTCTCTTGGGAAGAGGATATACATTTAGCCTTAAG 309  
 Db 970 GTGTG-----CCATCCAGTCCATGCGAGGCCGAGGAGCTCAGGTACAGCCTGCAT 1020  
 QY 310 GTTGACCATGGAGTAAATGCAGACTGCTCATCTTAAAGAAATTAATCCAAAGCGGAGA 369  
 Db 1021 GTGGGCTCTGGAGCACCCTGCTCAATGCCCACTCCGACAGTAAGACAGCAAGGAGA 1080  
 QY 370 ACTGTTCTGGATTTTAA-----CTCTGATTCAAATGAG 402  
 Db 1081 CGCGGAAAGATAAAGAACGGGAAAGACCGCAAGAGGAGTAAGAGATCCAGAAGCC 1140  
 QY 403 CGAGTCACCTTTAAACATCAAGTTTACA-----AAGCACATCATCTCCGAAGTCT 453  
 Db 1141 CCGAGCTTTATTAGAAAAGAGAAACAGAAACAGGAGGAGAACAGACAGAGAAACAATAT 1200  
 QY 454 TGGGCAATAGAGATAGGTTATCAAAACCGCGCAGGTTTCTGTGTACAAGAAGTATGGACAA 513  
 Db 1201 TGGGACATCCAGATTGGATATCAGACCAAGAGGTTATGTGTCATTAAACAAGCGGGGAAA 1260  
 QY 514 AATGCTATGTTAAGCCTTGGCTTCAAGATTCCTCCCATGACTGTTCAGTCTGATC 573  
 Db 1261 GCTGCTGATTTAAGCTTTTGGCAGCAAGAGAGCTTCAATGACCTTCCAGTCTGTGTG 1320  
 QY 574 ATGCCAAAGACTGTGAAACCTCCAGTGGTCTCTGGAGCCCTGTCTCAAGACATGC 633  
 Db 1321 ATCCAAAGAGTGCAGGTTTCCAGTGGTGGAGTGGAGCCCTGCTCAAAACATGC 1380

Direct Submission  
 Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7  
 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan  
 (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)  
 NEO human cDNA sequencing project supported by Ministry of  
 Economy, Trade and Industry of Japan; cDNA full insert sequencing:  
 Research Association for Biotechnology (RAB); cDNA library  
 construction: Helix Research Institute (HRI) (supported by Japan  
 Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,  
 HRI, and Biotechnology Center, National Institute of Technology and  
 Evaluation; clone selection for full insert sequencing: HRI and  
 RAB; annotation: HRI and RAB.  
 Location/Qualifiers  
 1. 3053  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="NT2RP7005669"  
 /cell\_line="NT2"  
 /cell\_type="teratocarcinoma"  
 /clone\_lib="NT2RP7"  
 /note="cloning vector: pME18SFL3-mRNA from NT2 neuronal precursor cells after 5-weeks retinoic acid (RA) induction."

QY 634 CGTTGAGGAGTCTCTTCCAGGATTTAGGAGCAGGAGCCGGAAACGTTGAAGCAATGGCT 693  
 Db 1381 CATGACATGGTGTCCCTTCCAGGACCTCTGTGAAGGACACGAACCATCAGGACGTTTCCC 1440  
 QY 694 ATTGAGGTGGAAAGAGTGTCTGAACTTCTTGAGAAAGAGGCTGCTATTTGTAAGGA 753  
 Db 1441 ATTGCACTGAAAGAGTGTCCAGAAATTTGAAGAAAGAAACCTGTTGTCTCAAGGA 1500  
 QY 754 GAACTTCTGAGCAATGTCACAGTATTCCTGGAGAACTTCTGAATGGAAGAAATGCCAA 813  
 Db 1501 GATGAGTGTGCTCCCTGTGCCACGATGCTGCTGGAACCTACAGAGTGTGCTGAGTGCCT 1560  
 QY 814 GTCCTCTCTCTCTCGACAGCAGGATCCCACTGGCATGTGAGGAGCCGCTGTGGC 873  
 Db 1561 GTGGACCTTTGCTCAGTCAGCAGCAGAGAGGCGCAACGAGCGGCCCTCTGTGGA 1620  
 QY 874 GGTGGGATCCAGACCCGGAGGTGTACTGTGCCAG-----AGCGTACCAGCA 921  
 Db 1621 GGGGCACTCCAGACCCGAGAGGTGTACTGCGTGAGGCCAACGAAACCTCTCTCAAA 1680  
 QY 922 GCTGCGCACTGAGGCCAAGGAGTCTCTAGACTGTGGAAGGCAATATGTGTGGA 981  
 Db 1681 TTAAGTACCCACAGAAACAAAGAGCTTCAAGCCATGGAATTAATATGCACTGGA 1740  
 QY 982 CCGGCCCGTTGCCCTCTCAGCTCTGCAATATCCCTTGTCTACGGACTGCAATATCT 1041  
 Db 1741 CTTATCCCTAATACTACACAGCTGTGCCACATCTCTTGTCCAACTGATGTGAATTTCA 1800  
 QY 1042 TCCGTGTGAGCTGGGCTGTGATCCATGAACTGTGTAAGCTTCAAGGAGGAAAAA 1101  
 Db 1801 CTTTGGTCAAGTGGGACCTTGTACTTATGAAACCTGTATGATGACGAGGAGAAAAA 1860  
 QY 1102 GGAATTTAGAACGAGCAGC-----CCATGTCTCTCATGGAATCTACAGGCCCTGCA 1152  
 Db 1861 GCTTCAACTGAGGAGCGCGGATACCAATGAGCCCTGAGGCTCTGGGCTAAC 1920  
 QY 1153 GGGCAATCCCTCAATTTGGTGGAGTGTCTTCTGTGAGGATCAATGTGTACCGATG 1212  
 Db 1921 GGAACCTGCCCTCACTTACTGGAAGCCATTCCTCTGGAAGGCTGCTGTATGACTGG 1980  
 QY 1213 CTGGCA---TCAGAGGAGTCTGTTCCTCTCATATGAA---AATGTGCTGGGACAT 1266  
 Db 1981 AAGCAGTGTGAGTGGGAACTGGAGCCAGATTAACGAAAGAGTGTGTCCAGGCAG 2040  
 QY 1267 CGTATTTGAAGGCGTCTGCCAATGACCGCGAGAGATGATPACAGGAGTCTTTGC 1326  
 Db 2041 CAAGTTCAAGAGGTGTGTGCATCAACAGTGTGAGAGAGTGTGACAGCAGCTGTGC 2100  
 QY 1327 CCAGTTCCCTCTCTCTGAGAGGAGTCTTGTGAATTCCTGCCAATGGAGTGTGTG 1386  
 Db 2101 AGAGATGCCATCTTCCCATCCCTGTGGCCGTGTATGCCCATGCCGAAAGACTGTGTG 2160  
 QY 1387 CTGAGCGAGTGGAGGAGTGTCTATCTGTGTTCCAGTCTGTTCATAAATAAAACTCAGAT 1446  
 Db 2161 CTCAGCAGATGTTACGTGTCTCTCTGTCTCACACACTGCTCAGGGAACGACAGAA 2220  
 QY 1447 GGGAAACAGACAGGTCAAGAACTATCTGTGCACTGTGG-----GAAGTGGAAAGCCCA 1503  
 Db 2221 GGGAAACAGATACAGAGCAGCATCTTCTGTGCTTATGCGGTGAGAGGAGTGTGCTGC 2280  
 QY 1504 TGTCCCTCTAGTCAAGGCTCTCCAGAGAGCATCTGTTGTGTATGACCATCTCTGTATGACG 1563  
 Db 2281 TGTCCAAATAGCAGTCTTTCAGAGAGTACGAGCTGTATGAGCATCTTGTGACAGTG 2340  
 QY 1564 CTTTCTGGGAGACATGCCCTTGGGCGCTTGTCTGAGGACATTTGGTAATGGCCCTT 1623  
 Db 2341 TACCCTGGCAACTGTGTCTTGGGCGCAGTGTGATTTGAGGACACTCAGTATCGTCTTC 2400  
 QY 1624 AATGCAACATTTGGCTGGAATGGAGAGCCAGCTGTGTGTAGGCACTTCAAGTCCGAGA 1683  
 Db 2401 AACAACACTACACTTGGATGGGAGGCTCTCTGCTGTGGCAGTGCAGCAAGAAAA 2460

QY 1684 GTCTTCTGTGTCAAGAGTCACTGGGACAAAGTAATGACCAAAAGATGTCAGATTTACT 1743  
 Db 2461 GTCATCTGTGTGCGAGTCAATGTGGGCCAAAGTGGGACCCCAAAAATGCTCTGAAAGCCTT 2520  
 QY 1744 CGACCTGAAACTGTGGCCCTGTTTCTCCCATGCAAAAAGACTGTATTTGTGCTGT 1803  
 Db 2521 CGACCTGAAACTGTGAAGCCTTGTCTGCTTCTTGTGAAGAGGACTGTATTTGTGACCCA 2580  
 QY 1804 TTAGTGTAGTGAACACCTTGGCCCAAGGATGTCCCAAGCAGGAAATGCCACAGT---AAAA 1860  
 Db 2581 TATAGTGTAGTGAACATCATGCCCCCTCTTGTGTAAAGAGGAGGACTCCAGTATCAGGAAG 2640  
 QY 1861 CAGTCTCGATACAGATCATCATCAAGAGCAGCAATGAGGCGAGGATGCCAGAT 1920  
 Db 2641 CAGTCTAGGCACTGGGTCTATCATTCAGCTGCCAGCAACGCGGGCGGAGACTGCCAGAT 2700  
 QY 1921 ACCTTATATGAGGAGAGAGTGTGAAGATGTTCTTGTGTCTCTGTATATATCGTGGAG 1980  
 Db 2701 CCCCTCTATGAAGAGAGGCTGTGAGCACCCTCAAGCGTGCCTCAAGCTTACAGGTGGAAG 2760  
 QY 1981 CCACAGAAATGAGGCGCTTGTGATCTTAGTGCCAGAGTCTGTCTGCGAGGAAATAACGGGC 2040  
 Db 2761 ACTCAAAATGCGCGAGATGCCAATTAGTCCCTTGGAGCGTGCACAAAGACAGAGCCCTGGA 2820  
 QY 2041 AGCAGTGAAGCCTGTGGAAAGGGTTACAAACAGAGCTGTCTCATGATCTCTCTGATGAC 2100  
 Db 2821 GCACAGGAGGCTGTGGGCTTGGCGCAGCAGCAAGGCGCATTTCTTGTGCAAGCAAGAT 2880  
 QY 2101 AACCGTTCAGCAAAATGATGAATGCTTCAAGCAGACAAACGGGATGCTCTCTCTGTG 2160  
 Db 2881 GGAGGACAGGCTGGAATCCATGATGCTCAGTATGAGGCGCTGTGCCAGCCCTTACC 2940  
 QY 2161 CAAGAATGACAGTCCCATGTGCGAAGAGTGCACCTTCACTGTCTGCTTGTGCTCAAGTTTACG 2220  
 Db 2941 CAGGCTGCGAGATCCCTGCGAGAGTACTGTCAATTGACAGCTGTGCCAAGTTTCT 3000  
 QY 2221 CCTGTCTCCAGAACTGTGAAGCCACAAAGTAGTGGCGCG 2261  
 Db 3001 TCATGCAATGGAGACTGTGTGCTAGTCTAGGACCAAGAGCG 3041

RESULT 12  
 AC011231/10  
 LOCUS Homo sapiens BAC clone RP11-13K3 from 2, complete sequence.  
 DEFINITION AC011231  
 AC011231.7 GI:14190752  
 VERSION HTG.  
 KEYWORDS Homo sapiens (human)  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 119205)  
 Sultston, J.E. and Waterston, R.  
 Toward a complete human genome sequence  
 Genome Res. 8 (11), 1097-1108 (1998)  
 99063792  
 9847074  
 REFERENCE 2 (bases 1 to 119205)  
 Harris, A. and Maupin, R.  
 The sequence of Homo sapiens BAC clone RP11-13K3  
 Unpublished  
 3 (bases 1 to 119205)  
 Waterston, R.H.  
 Direct Submission  
 Submitted (04-OCT-1999) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 4 (bases 1 to 119205)  
 Waterston, R.H.  
 Direct Submission  
 Submitted (23-MAY-2001) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA  
5 (bases 1 to 119205)  
Waterston, R.  
Direct Submission  
Submitted (07-NOV-2001) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On May 23, 2001 this sequence version replaced gi:12431080.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)  
----- Summary Statistics  
Center project name: H NH00113K03

**NOTICE:** This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

**MAPPING INFORMATION:**  
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

**SOURCE INFORMATION:**

The RP61-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P. Y., Zhao, B., Frengen, E., Iateno, M., Catanesi, J. and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

NEIGHBORING SEQUENCE INFORMATION:  
The clone sequenced to the left is RP11-398K16, 200 bp overlap; the clone sequenced to the right is RP11-112N16. Actual start of this clone is at base position 138700 of RP11-398K16; actual end is at base position 7193 of RP11-112N16

FEATURES	source
base position / 1993 or RP11-112N1	Location/Qualifiers
	1. .119205
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="2"
	/map_pos="2"
	/clone="RP11-13K3"
	/clone_lib="RP11-11"
	2. .164
repeat_region	/rpt_family="L2"
repeat_region	7728. .7785
repeat_region	/rpt_family="L2"
repeat_region	7805. .8116
repeat_region	/rpt_family="Alu"
repeat_region	8482. .8544
repeat_region	/rpt_family="L2"
repeat_region	8671. .8748
repeat_region	/rpt_family="Mariner"
repeat_region	10478. .10853
repeat_region	/rpt_family="L2"
repeat_region	11336. .11375
repeat_region	/rpt_family="(TG)n"

```

11737. .11771
/rpt_family="AT_rich"
repeat_region
12229. .12423
/rpt_family="MIR"
repeat_region
13139. .13436
/rpt_family="Alu"
repeat_region
13554. .13694
/rpt_family="MIR"
repeat_region
13704. .13718
/rpt_family="MIR"
repeat_region
13719. .13765
/rpt_family="L2"
repeat_region
13766. .13892
/rpt_family="MIR"
misc_feature
14019. .14272
[note="similar to EST BE299433 (NID:99183181)"]
repeat_region
14575. .14855
/rpt_family="Alu"
repeat_region
14903. .15059
/rpt_family="MIR"
repeat_region
15247. .15381
/rpt_family="L1"
repeat_region
15621. .15826
/rpt_family="MER1_type"
repeat_region
16211. .16507
/rpt_family="Alu"
repeat_region
16695. .16988
/rpt_family="Alu"
repeat_region
17012. .17049
/rpt_family="AT_rich"
repeat_region
17297. .17621
/rpt_family="MaLR"
repeat_region
17622. .17653
/rpt_family="(TAAAA)n"
repeat_region
17654. .17715
/rpt_family="MaLR"
repeat_region
19019. .19049
/rpt_family="AT_rich"
repeat_region
19050. .19108
/rpt_family="L1"
repeat_region
19109. .19386
/rpt_family="Alu"
repeat_region
19387. .19705
/rpt_family="L1"
repeat_region
19706. .19870
/rpt_family="(TA)n"
repeat_region
19890. .20882
/rpt_family="L1"
repeat_region
20883. .21186
/rpt_family="Alu"
repeat_region
21187. .21558
/rpt_family="L1"
repeat_region
21559. .21583
/rpt_family="(TTTGT)n"
repeat_region
21584. .21872
/rpt_family="L1"
repeat_region
21950. .21984
/rpt_family="AT_rich"
repeat_region
22011. .22261
/rpt_family="L1"
repeat_region
22474. .22706
/rpt_family="MaLR"
repeat_region
22814. .23079
/rpt_family="Alu"
repeat_region
23190. .23461
/rpt_family="L1"
repeat_region
23518. .24159
/rpt_family="L1"
repeat_region
24270. .24652
/rpt_family="MaLR"
repeat_region
24696. .24959
/rpt_family="L1"
repeat_region
26424. .26590

```

```

/rpt_family="MIR"
27884. .28177
/rpt_family="L1"
28632. .28930
/rpt_family="Alu"
28940. .28990
/rpt_family="CATATA,n"
29005. .29117
/rpt_family="Alu"
29554. .29981
/rpt_family="L1"
29982. .30099
/rpt_family="Alu"
30100. .30583
/rpt_family="L1"
30826. .32314
/rpt_family="L1"
38581. .38608
/rpt_family="AT-rich"
38609. .38984
/rpt_family="L1"
39073. .39148
/rpt_family="L1"
39214. .39340

Query Match
Best Local Similarity 100.0%; Score 527; DB 9; Length 119205;
Matches 527; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTCGCGCTGATTCAGAGCTGAAACCGAAGCTGTTGTTGCAATGAAATATGCGAACAC 60
DB 52441 ATGTCGCGCTGATTCAGAGCTGAAACCGAAGCTGTTGTTGCAATGAAATATGCGAACAC 52382
QY 61 TTGTCGCTTACGCTCCCTACAGACAGGCTTGCCTCATCTCTGTCGCCGGATTTGTTGA 120
DB 52381 TTGTCGCTTACGCTCCCTACAGACAGGCTTGCCTCATCTCTGTCGCCGGATTTGTTGA 52322
QY 121 GTATCTGAGTTCTTACCATGTCCTGTCAGTGTGAGCAAGGATGTGGAGAAATTCAGCAT 180
DB 52321 GTATCTGAGTTCTTACCATGTCCTGTCAGTGTGAGCAAGGATGTGGAGAAATTCAGCAT 52262
QY 181 AGAATCGCGGGTCATAGCTCCCTCTCTTTGTTGTTGCTTTCGCAATGTCGCAATCTGACT 240
DB 52261 AGAATCGCGGGTCATAGCTCCCTCTCTTTGTTGTTGCTTTCGCAATGTCGCAATCTGACT 52202
QY 241 GAGTCAAGAGCCTGTGATGCTCCCATTTCTGCTCTCTTGGGGAAGGGAATATACATTT 300
DB 52201 GAGTCAAGAGCCTGTGATGCTCCCATTTCTGCTCTCTTGGGGAAGGGAATATACATTT 52142
QY 301 AGCCTTAGGTTGGACCATGAGTAAATGACAGCTGCTGCTGCTTAAAGAAATTAATCCA 360
DB 52141 AGCCTTAGGTTGGACCATGAGTAAATGACAGCTGCTGCTGCTTAAAGAAATTAATCCA 52082
QY 361 AGCGGAAGAACTGTTCTCGAATTTAACTCTGATTCATGAGCGAGTCCACCTTTAAACAT 420
DB 52081 AGCGGAAGAACTGTTCTCGAATTTAACTCTGATTCATGAGCGAGTCCACCTTTAAACAT 52022
QY 421 CAAAGTTCAAAGACATCATCATTCGAAGTCTTGGCAATAGAGATAGGTTATCAACC 480
DB 52021 CAAAGTTCAAAGACATCATCATTCGAAGTCTTGGCAATAGAGATAGGTTATCAACC 51962
QY 481 CGCGAGGTTTCGTGTACAGAGAGTATGAGCAAAATGCTATGTTAAG 527
DB 51961 CGCGAGGTTTCGTGTACAGAGAGTATGAGCAAAATGCTATGTTAAG 51915

RESULT 13
AC027600/c 167462 bp DNA linear HTG 07-JUN-2000
LOCUS Homo sapiens chromosome 2 clone RP11-41418 map 2, WORKING DRAFT
DEFINITION SEQUENCE, 22 unordered pieces.
ACCESSION AC027600
VERSION AC027600.2 GI:8318467

```

KEYWORDS  
SOURCE  
ORGANISM

HTG; HTGS PHASE1; HTGS DRAFT.  
Homo sapiens (human)  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 167462)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
Homo sapiens chromosome 2, clone RP11-41418  
Unpublished

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

1 (bases 1 to 167462)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,  
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,  
Campopiano,A., Castle,A., Chospel,Y., Colangelo,M., Collins,S.,  
Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., D.,  
Dodge,S., Domino,M., Doyle,M., Ginde,S., Goyette,M., Gage,D.,  
Galan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczkv,J.,  
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,  
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,  
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,  
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
O'Neill,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,  
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
Stange,Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,  
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
Young,G., Zainou,J., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (30-MAR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jun 7, 2000 this sequence version replaced gi:7342347.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE  
JOURNAL  
COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L9157  
Center clone name: 414.1.8  
----- Summary Statistics  
Sequencing vector: M13; M7815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 156843 bases at least Q40  
Consensus quality: 161829 bases at least Q30  
Consensus quality: 163971 bases at least Q20  
Insert size: 170000; agarose-fp  
Insert size: 163362; sum-of-contigs  
Quality coverage: 4.2 in Q20 bases; agarose-fp  
Quality coverage: 4.3 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 22 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 1293: contig of 1293 bp in length  
\* 1294 1393: gap of 100 bp  
\* 1394 3228: contig of 1835 bp in length  
\* 3229 3328: gap of 100 bp  
\* 3329 5140: contig of 1812 bp in length  
\* 5141 5240: gap of 100 bp  
\* 5241 8017: contig of 2777 bp in length



Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE  
AUTHORS

1 (bases 1 to 224789)  
Munz D Marie, Metaker M Lee, Abranzon S, Adams C, Alder J, Allen C, Allen H, Albrooks S, Amin A, Anguiano D, Anyalebechi V, Aoyagi A, Ayodeji A, Bacca E, Baden H, Baldwin D, Bandaranaike D, Barber M, Barnstead M, Benahmed F, Biswal K, Blair J, Blankenburg K, Blyth P, Brown M, Bryant N, Buhay C, Burch P, Burrell K, Calderon E, Cardenas V, Carter K, Cavazos I, Ceasar H, Chen A, Chacko J, Chavez D, Chen G, Chen R, Chen Y, Chen Z, Chu J, Cleveland C, Cockrell R, Cox C, Coyle M, Cree A, D'Souza B, Davila M, Davis C, Davy-Carroll J, De Anda C, Dederich D, Delgado O, Denison S, Deramo C, Ding Y, Dinh H, Divya K, Draper H, Dugan-Rocha S, Dunn A, Durbin K, Duval B, Eaves K, Egan A, Escotto M, Eugene C, Evans C, Falls T, Fan G, Fernandez S, Finley M, Flagg N, Forbes L, Foster M, Foster P, Fraser C, Gabisi A, Ganta R, Garcia A, Garner T, Garza M, Gregoridis E, Geer K, Gill R, Grady M, Guerra W, Guevara W, Guaratone P, Haaland M, Hamill C, Hamilton C, Hamilton K, Harvey I, Havlak P, Hawes A, Henderson N, Hernandez J, Hines S, Hladun S, Hodgson A, Hogues M, Hollins B, Howells S, Hulyk S, Hume J, Idlebird D, Jackson A, Jackson L, Jacob L, Jiang H, Johnson B, Johnson R, Jolivet A, Karpathy S, Kelly S, Kelly S, Khan Z, King L, Kovar C, Kowis C, Kraft C, Lebow H, Levan J, Lewis L, Li Z, Liu J, Liu J, Liu W, Liu Y, London P, Longacre S, Lopez J, Lorenshewa L, Loulseghe H, Lozano R, Lu X, Ma J, Maheshwari M, Mahindartne M, Mahmoud M, Malloy K, Mangum A, Mapua P, Martin K, Martin R, Martinez E, McWhiney S, McLeod M, McNeill T, Meenen E, Milosavljevic A, Miner G, Minja E, Montemayor J, Moore S, Morgan M, Morris K, Morris S, Munidasa M, Murphy M, Nair L, Nankervis C, Neal D, Newton N, Nguyen N, Norris S, Nwankwelu O, Okwou G, Olaprunsaagoo A, Pal S, Parks K, Pasternak S, Paul H, Perez A, Perez L, Pfannkuch C, Plopper F, Poindecker A, Popovic D, Primus E, Pu L, Pu L, Puazo M, Quiroz J, Rachin E, Reeves K, Regier M, Reigh R, Reilly B, Reilly M, Ren Y, Reuter M, Richards S, Riggs P, Rives C, Rodkey T, Rojas A, Rose M, Rose R, Ruiz S, Sanders W, Savary G, Scherer S, Scott G, Shatsman S, Shen H, Shetty J, Shvartsbeyn A, Sisson I, Sitter C, Smajls D, Sneed A, Sodergren E, Song X, Sorelle R, Sosa J, Steime M, Strong R, Sutton A, Svatek A, Taber P, Taylor C, Taylor T, Thomas N, Thomas S, Tingey A, Trejos Z, Usmani K, Valas R, Vera V, Villanasa D, Waldron L, Walker B, Wang J, Wang Q, Wang S, Warren J, Warren R, Wei X, White F, Williams G, Wilson R, Wleczky R, Wooden H, Worley K, Wright D, Wright R, Wu J, Yakub S, Yen J, Yoon L, Yoon V, Yu F, Zhang J, Zhou J, Zhou X, Zhao S, Dunn D, von Niederhauser A, Weiss R, Smith D, Holt R, Smith H, Weinstock G, and Gibbs R.A.

TITLE  
JOURNAL

2 (bases 1 to 224789)  
Worley K.C.  
Direct Submission  
Submitted (19-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE  
AUTHORS

3 (bases 1 to 224789)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (13-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

TITLE  
JOURNAL

On Nov 13, 2002 this sequence version replaced gi:22857292.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold,

individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: GMSK  
Center clone name: CH230-237P17  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 209106 bases at least Q40  
Consensus quality: 213906 bases at least Q30  
Consensus quality: 216832 bases at least Q20  
Estimated insert size: 218750; sum-of-contigs estimation  
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 1 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 224789: contig of 224789 bp in length.

FEATURES  
source

1. 224789  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-237P17"

ORIGIN

Query Match 8.2%; Score 362.4; DB 2; Length 224789;  
Best Local Similarity 80.7%; Pred. No. 4.4e-98;  
Matches 423; Conservative 0; Mismatches 101; Indels 0; Gaps 0;  
QY 4 GTGCGCTGCAATTCAGAGCTGACCGAAGTGTGTTGCAATGAAATATGCGAACACTTT 63  
Db 43659 GTGCGCTGCTTTCAAGAGCTCAACAGAACCAATGTTGGCCAGTGAGATCTGTGAACACTTT 43718  
QY 64 GCCCTTCAGCTTCCTACAGAACAGAGCTTGCCTCATCTCTTGTCCCGGATGTGTAGTA 123  
Db 43719 GTTCTCAGCTCCACAGAGCAAGCTTGCTCATCTCTTGTCCCGGATGTGTAGTA 43778  
QY 124 TTGAGTTCTTACCATGTCCTCACTGTAGCAAGGATGTGGAGAGAAATTCAGCATAGA 183  
Db 43779 TCTGAGTTCTCCCGCTGTGTCACCTGCTGCTCATCTCTTGTCCCGGATGTGTAGTA 43838  
QY 184 ACTCGCGGCTCATAGCTCCCTCTCTTTGTGTGTTGCAATGTCCTCAATCTGACTGAG 243  
Db 43839 ACTCGGTGCGCATCGACCCCTCTGTATGAGAGTCTGAGTGTGCTGCTGCTGCTGAG 43898  
QY 244 TCAAGAGCTGTGATGCTGCCATTTCTCTCTCTTGTGGGAGAGAGAAATATACATTTAGC 303  
Db 43899 TCCAGAGCTGTGAGGCTCCCAATTTCTCTCTCTTGTGGGAGAGAGAAATATTTCTTCAGC 43958  
QY 304 CTTAAGTTTGCACCTGAGTAATGCGAGCTGCTCATCTTAAAGAAATTAATCCAGC 363  
Db 43959 CTTAAGTTTGCACCTGAGTAATGCGAGCTGCTCATCTTAAAGAAATTAATCCAGC 44018  
QY 364 GGAAGAACTGTTCTGGATTTTAACTCTCTGATTCAAATGAGCGAGTCACCTTTAAACATCAA 423



Db 44019 GGAAGAAATGTCAGGATTTAGCTCTGATTCAACGAGCAAGTCACCCCTAACACATCAG 44078

QY 424 AGTTACAAAGCACATCATCTTCGAGCTTTGGCAATAGAGATAGATTATCAACCCCG 483

Db 44079 AGTTACAAAGCACAGCACCTCTCCAGCTTGGGCTATAGAGTAGTTTCCAAACAGG 44138

QY 484 CAGGTTCTGTTACAAAGATGATGGACAAATATGCTATGTAAG 527

Db 44139 CAGGTGTGGTGCACAGAGTGTGGAGAAAGCCCTGTGTAAG 44182

RESULT 15

AC123741/c

LOCUS AC123741.4 GI:30017756

DEFINITION HTG; HTGS PHASE1; HTGS DRAFT.

VERSION AC123741

KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 259236)

AUTHORS Birren, B., Nusbaum, C. and Lander, E.

TITLE Mus musculus, clone RP24-113A5

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 259236)

AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, V., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzgerald, M., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE Direct Submission

JOURNAL Submitted (01-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 3 (bases 1 to 259236)

AUTHORS Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Canarata, J., Chang, J., Choepel, V., Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,

TITLE  
JOURNAL

## COMMENT

Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (23-APR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Apr 17, 2003 this sequence version replaced gi:28882446.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
-----  
Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBX  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
-----  
Project Information  
Center project name: L26529  
Center clone name: 113\_A\_5  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 5 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 655: contig of 655 bp in length  
\* 656 755: gap of 100 bp  
\* 756 1434: contig of 679 bp in length  
\* 1435 1534: gap of 100 bp  
\* 1535 7687: contig of 6153 bp in length  
\* 7688 7787: gap of 100 bp  
\* 7788 83305: contig of 75518 bp in length  
\* 83306 83406: gap of 100 bp  
\* 83406 259236: contig of 175831 bp in length.

FEATURES  
source

1. 259236  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/clone="RP24-113A5"  
/clone\_lib="RPC1-24 Male Mouse BAC"

## misc\_feature

1. 655  
/note="assembly\_fragment"

## misc\_feature

756..1434  
/note="assembly\_fragment"

## misc\_feature

1535..7687  
/note="assembly\_fragment"

## misc\_feature

7788..83305  
/note="assembly\_fragment"

## misc\_feature

83406..259236  
/note="assembly\_fragment"

clone\_end:T7  
vector\_side:right

## ORIGIN

Query Match 8.2%; Score 360.8; DB 2; Length 259236;  
Best Local Similarity 80.5%; Pred. No. 1.4e-97;  
Matches 422; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 4 GTGGCTGCATTCAGAGCTGAACGAACTGTGTTGCAATGAATAATGGAACACTTT 63

Db 218506 GTGGCTGTCTTTCAGAGCTCAATAGAACCATGGTGTCCAATGAGATCTGTGAACACTTT 218447

QY 64 GCCCTTACGCTCTCAGAACAGCGCTTGCTCATTCCTTGTCGCCGGGATTGTGTAGTA 123

Db 218446 GCTCTCTAGCCCCCAGAGCAAGCCTGCTCTATTCTTGTCCCCCAGGAGCTGCTGTGTA 218387

QY 124 TCTGAGTTCTTACCATGCTCAACTGTAGCAAGGATGTGGAGAAATTCGACATAGA 183

Db 218386 TCTGAGTTCTTCCCATGCTGCACCTGCGCCGAGAGTGTGGAGAAATTCGACATAGA 218327

QY 184 ACTGCGCGGTATAGCTCCCCCTCTTTGGTGGTTTGCATGTCCCAATCTCAGTCAG 243

Db 218326 ACTCGTGTGGCCATCGGCCCCCTCTGTATGGAGGTCTGCAGTGTCAAATCTCAGGAG 218267  
Qy 244 TCAAGAGCCTGTGATGCTCCCAATTCCCTGCTCTTGGGGAAGAGGAATATACATTTAGC 303  
Db 218266 TCCAGAGCCTCGGAGGCTCCAGTTTCTGTCTCTTGGGAAAGAGAAATATTCCTTCAGC 218207  
Qy 304 CTTAAGGTTGGACCATGGAGTAAATGCAGACTGCTCATCTTAAAGNAATTAATCCAAGC 363  
Db 218206 CTAAGGTTGGACCATGGAGCAATGTAGACTCCCTCACCTTAAGGAAGTCGACCTCAGC 218147  
Qy 364 GGAAGAACTGTCTGGATTTTAACTCTGATTCAAATGAGCGAGTCACTTTAAACATCAA 423  
Db 218146 GCAAGAAATATCCAGGATTTTAGTCTGTGATTCRAACGAGCAAGTCACCTTAACACATCAG 218087  
Qy 424 AGTTACAAGCACATCATTCGAACTCTTGGCAATAGAGTAGGTTATCAAAACCCGG 483  
Db 218086 AGTTACAAGCACACCCACCACTCCAGCCTGGGGATGTAGTAGGGTTTCAAAACCCAGG 218027  
Qy 484 CAGGTTTCGTGTACAAGAAGTGTGACAAAAATGCTATGTTAAG 527  
Db 218026 CAGGTGTGGTGTACAGAAAGTGTGGAAGAAACGCCCTGTTAAG 217983

Search completed: February 23, 2004, 17:26:01  
Job time : 16550.2 secs

result No.	Score	Query		Length	DB	ID	Description.
		Match	%				
1	4380.4	99.6	5847	9	ADB07142	Novel	Cod
2	2335.8	50.8	3382	6	AA143703	Human	lar
3	1323.8	28.0	4810	7	ACC72174	Human	NOV
4	1332.8	28.0	6317	6	ACC728694	Human	lf2
5	1322.8	28.0	6373	4	AAR30188	Clone	214
6	1323.8	28.0	6373	7	ACC72162	Human	NOV
7	1332.8	28.0	6373	7	ACC72171	Human	NOV
8	1323.8	28.0	6373	7	ACC72168	Human	NOV
9	1329.6	28.0	6378	4	AA158451	Human	pol
10	1329.6	28.0	6378	8	ADB48422	Novel	hum
11	1190.4	27.1	6292	4	AA160237	Human	pol
12	1177.4	26.8	4847	7	ACC72167	Human	pol
13	1177.4	26.8	4881	7	ACC72165	Human	NOV
14	1333.4	25.8	4679	7	ACC72166	Human	NOV
15	580.6	13.2	3053	9	ADB62804	Human	CDN
16	541.4	12.3	1713	7	ACC72164	Human	NOV
17	540.2	12.3	1732	7	ACC72172	Human	NOV
18	512.4	11.7	516	6	ABK70149	Human	lun
19	505.8	11.5	1228	5	AA869250	DNA	encod
20	460	10.5	744	4	AAK43667	Human	bon
21	460	10.5	744	4	AAK17814	Human	bra
22	436.6	9.9	456	5	AA869249	Human	NOV
23	402.8	9.2	1401	7	ACC72163	Human	NOV

CC proteins. The DNA and protein sequences of the invention are useful as:  
CC markers for tissues in which the corresponding protein is preferentially  
CC expressed; as molecular weight markers on gels; as chromosome markers or  
CC tags; to identify chromosomes or to map related gene positions; and to  
CC compare with endogenous DNA sequences in patients to identify potential  
CC genetic disorders. The present DNA sequence represents a gene of the  
CC invention.

XX	Sequence	5847 BP; 1607 A; 1282 C; 1453 G; 1505 T; 0 U; 0 Other;
SQ	Query Match	99.6%; Score 4380.4; DB 9; Length 5847;
	Best Local Similarity	99.8%; Pred. No. 0;
	Matches 4397; Conservative	0; Mismatches 1; Indels 6; Gaps 1;
QY	1	ATGGTGCCTGATTCAGAACTGACCGAATCTGGTTGCAAAATGAAATATGCGAACAC 60
DB	322	ATGGTGCCTGATTCAGAACTGACCGAATCTGGTTGCAAAATGAAATATGCGAACAC 381
QY	61	TTTGCCCTTCAGCTCCCTACAGAACAGGCTTGCCTCATTCCTTCCCGGATTTGTGA 120
DB	382	TTTGCCCTTCAGCTCCCTACAGAACAGGCTTGCCTCATTCCTTCCCGGATTTGTGA 441
QY	121	GTATCTGAGTTCTTACCATGTCCTCAACCTGTAGCAAGGATGTGGAAAGAAATTCAGCAT 180
DB	442	GTATCTGAGTTCTTACCATGTCCTCAACCTGTAGCAAGGATGTGGAAAGAAATTCAGCAT 501
QY	181	AGAACTCCGCGGTGATAGCTCCCTCTCTTTGGTGGTTGCAATCTCCAAATCTGACT 240
DB	502	AGAACTCCGCGGTGATAGCTCCCTCTCTTTGGTGGTTGCAATCTCCAAATCTGACT 561
QY	241	GAGTCAAGAGCCTGTGATGCTCCATTTCTGCTCTCTTGGGGAAGAGGAATATACATTT 300
DB	562	GAGTCAAGAGCCTGTGATGCTCCATTTCTGCTCTCTTGGGGAAGAGGAATATACATTT 621
QY	301	AGCCTTAGGTTGGACCATGAGTAATGACAGTGGCTCATCTTAAGAAATTAATCCA 360
DB	622	AGCCTTAGGTTGGACCATGAGTAATGACAGTGGCTCATCTTAAGAAATTAATCCA 681
QY	361	AGCGAAGAACTGTTCTGGAATTTAACTCTGATTCAAATGAGCGAGTCACCTTTAAACAT 420
DB	682	AGCGAAGAACTGTTCTGGAATTTAACTCTGATTCAAATGAGCGAGTCACCTTTAAACAT 741
QY	421	CAAAAGTTCAAGACATCATCATTCGAAATCTTGGGCAATAGAGATAGGTATCAAAAC 480
DB	742	CAAAAGTTCAAGACATCATCATTCGAAATCTTGGGCAATAGAGATAGGTATCAAAAC 801
QY	481	CGGAGAGTTTCGTGTACAGAAAGTATGACAAAATGCTATGTTAAGCTTTTGCTTCAA 540
DB	802	CGGAGAGTTTCGTGTACAGAAAGTATGACAAAATGCTATGTTAAGCTTTTGCTTCAA 861
QY	541	GATTCTTCCCATTTGACCTGCTGATGCTGATGATGCTGATGCTGATGCTGATGCTGAT 600
DB	862	GATTCTTCCCATTTGACCTGCTGATGCTGATGATGCTGATGCTGATGCTGATGCTGAT 921
QY	601	TGGTCTCTCTGAGCCCTGCTCCAGACATGCTGCTTCCAGGAGTCTCTTCCAGAGATT 660
DB	922	TGGTCTCTCTGAGCCCTGCTCCAGACATGCTGCTTCCAGGAGTCTCTTCCAGAGATT 981
QY	661	AGGAGCAGAGCCGGAACGTGAAGACATGCTATGGAAGTGAAGAGGATGTCCTGAA 720
DB	982	AGGAGCAGAGCCGGAACGTGAAGACATGCTATGGAAGTGAAGAGGATGTCCTGAA 1041
QY	721	CTTCTTGGAAGAGGCTGCTGATTTGAGGAGAACTTCTGAGCAATGCTCCAGGTAT 780
DB	1042	CTTCTTGGAAGAGGCTGCTGATTTGAGGAGAACTTCTGAGCAATGCTCCAGGTAT 1101
QY	781	TCCTGGAGAACTTCTGATGAAAGAAATGCCAAGTCTCTCTCTCTCTGAGCAGCAGGAT 840
DB	1102	TCCTGGAGAACTTCTGATGAAAGAAATGCCAAGTCTCTCTCTCTCTGAGCAGCAGGAT 1161
QY	841	CCCCACTGGCATCTGACCGGACCCGCTGTGCGGTGGATCCAGACCCCGGAGGTGTAC 900
DB	1162	CCCCACTGGCATCTGACCGGACCCGCTGTGCGGTGGATCCAGACCCCGGAGGTGTAC 1221

QY	901	TGTGCCAGAGCTACCGAGCTGCCGACCTAGAGGCGCAAGGAAGTCTCTAGACCTGTG 960
DB	1222	TGTGCCAGAGCTACCGAGCTGCCGACCTAGAGGCGCAAGGAAGTCTCTAGACCTGTG 1281
QY	961	GAAAGGCATTTATGTGTGGGACCCCGCGTGGTCCCTCTCAGCTCTGCAATATCCCTTGC 1020
DB	1282	GAAAGGCATTTATGTGTGGGACCCCGCGTGGTCCCTCTCAGCTCTGCAATATCCCTTGC 1341
QY	1021	TCTACGAGCTGATAGTATCTTCCCTGCTCAGCTGGGCTGTGATCCATGAAACTGT 1080
DB	1342	TCTACGAGCTGATAGTATCTTCCCTGCTCAGCTGGGCTGTGATCCATGAAACTGT 1401
QY	1081	CATGAACCTCAGGGGAAATAAGGATTTAGACAGGCGGCGCATGTCTCATGAAATCT 1140
DB	1402	CATGAACCTCAGGGGAAATAAGGATTTAGACAGGCGGCGCATGTCTCATGAAATCT 1461
QY	1141	ACAGGCGCTGACAGGCAATTCCTCATTTGGTGGAGTCTGTTCCCTTGTGAGGATCAATG 1200
DB	1462	ACAGGCGCTGACAGGCAATTCCTCATTTGGTGGAGTCTGTTCCCTTGTGAGGATCAATG 1521
QY	1201	TGCTACCGATGGCTGGCATCAGAGGATCTGTTTCCCTGATCATTGGAAATGTGGCCTG 1260
DB	1522	TGCTACCGATGGCTGGCATCAGAGGATCTGTTTCCCTGATCATTGGAAATGTGGCCTG 1581
QY	1261	GGACATCTGTTCTCAAGCGCTGCGCAGATGACCGCGGAGAGATGTATCAGGAGT 1320
DB	1582	GGACATCTGTTCTCAAGCGCTGCGCAGATGACCGCGGAGAGATGTATCAGGAGT 1641
QY	1321	CTTTGCCAGTTCCCTCTCTCTCAGAGGAGTCTTGTGAAATTTCCCTGCCGATGGAC 1380
DB	1642	CTTTGCCAGTTCCCTCTCTCTCAGAGGAGTCTTGTGAAATTTCCCTGCCGATGGAC 1701
QY	1381	TGCTGTCTGAGCGAGTGGACGAGTGTCTCTCTGTTCCAGTCTCTTCAATAAATAAC 1440
DB	1702	TGCTGTCTGAGCGAGTGGACGAGTGTCTCTCTGTTCCAGTCTCTTCAATAAATAAC 1761
QY	1441	TCAGATGGGAAACAGACACAGGCTCAAGAACTATCTCTGCACTGGCTGGGGAAGTGGAAAG 1500
DB	1762	TCAGATGGGAAACAGACACAGGCTCAAGAACTATCTCTGCACTGGCTGGGGAAGTGGAAAG 1821
QY	1501	CCATGTCCCTTAGTCAGGCTCTCAAGAGCATCTGTTGTGTATGACCATCTCTGTATG 1560
DB	1822	CCATGTCCCTTAGTCAGGCTCTCAAGAGCATCTGTTGTGTATGACCATCTCTGTATG 1881
QY	1561	CAGCTTCACTGGGAGACATCGCTTTGGGGCCCTTCTCTGAGGACACATTTGGTAATGCC 1620
DB	1882	CAGCTTCACTGGGAGACATCGCTTTGGGGCCCTTCTCTGAGGACACATTTGGTAATGCC 1941
QY	1621	CTTAATGCAACCATTTGGCTGGAAATGGAGAACCCAGTGTGTGTAGGCATTCAGACTGG 1680
DB	1942	CTTAATGCAACCATTTGGCTGGAAATGGAGAACCCAGTGTGTGTAGGCATTCAGACTGG 2001
QY	1681	AGAGTCTTCTGTCTCAAGAGTCACTGGGGAACAAGTAATGACCAAAAGATGTCCAGATTCT 1740
DB	2002	AGAGTCTTCTGTCTCAAGAGTCACTGGGGAACAAGTAATGACCAAAAGATGTCCAGATTCT 2061
QY	1741	ACTGCACTGAAACTGTGTGGCCCTCTGTTTCTCCCATGCAAAAAGACTGTATTGTGACT 1800
DB	2062	ACTGCACTGAAACTGTGTGGCCCTCTGTTTCTCCCATGCAAAAAGACTGTATTGTGACT 2121
QY	1801	GCCTTTCAGTGTGACACCTGCGCCAGGATGTGCCAAGCAGGAAATGCCACAGTAAAA 1860
DB	2122	GCCTTTCAGTGTGACACCTGCGCCAGGATGTGCCAAGCAGGAAATGCCACAGTAAAA 2181
QY	1861	CAGTCTCGATACAGAAATCATCATCAAGAACAGCCTAATGGAGGCGCAGGAATGCCAGAT 1920
DB	2182	CAGTCTCGATACAGAAATCATCATCAAGAACAGCCTAATGGAGGCGCAGGAATGCCAGAT 2241
QY	1921	ACCTTATATGAGGAGAGAGTGTGAAGATGTTTCTTGTCTCTGTATATCGGTGGAG 1980
DB	2242	ACCTTATATGAGGAGAGAGTGTGAAGATGTTTCTTGTCTCTGTATATCGGTGGAG 2301

1981 CCACAGAAATGGAGCCCTTGCATCTTAGTSCCAGAGTCTGTCTGCGAGGAATAACGGGC 2040  
1982 CCACAGAAATGGAGCCCTTGCATCTTAGTSCCAGAGTCTGTCTGCGAGGAATAACGGGC 2041  
2041 AGCAGTGAAGCCCTTGGAAAGGGTTACAAACAGAGCTGTCTCATGCACTCTCTGATGAC 2100  
2362 AGCAGTGAAGCCCTTGGAAAGGGTTACAAACAGAGCTGTCTCATGCACTCTCTGATGAC 2421  
2101 AACCGGTCAAGCAAAATGATGAATGCCCTCAAGCAGACAAACGGCATGCCCTCTCTG 2160  
2422 AACCGGTCAAGCAAAATGATGAATGCCCTCAAGCAGACAAACGGCATGCCCTCTCTG 2481  
2161 CAAGATGACAGCTCCATGTGAGAGAGCTGACCTTCACTTGTGTCGAAGTTAG 2220  
2482 CAAGATGACAGCTCCATGTGAGAGAGCTGACCTTCACTTGTGTCGAAGTTAG 2541  
2221 CCTGCTCCACGAACCTGTGAAGCCACAAAAGTAGGCGCGCAGCTACAGCGGAAAGC 2280  
2542 CCTGCTCCACGAACCTGTGAAGCCACAAAAGTAGGCGCGCAGCTACAGCGGAAAGC 2601  
2281 AGAAGAGAGGAATGCCAGGATTTGACCTTTACCTTCTAGTGGAGACAGAACTATGT 2340  
2602 AGAAGAGAGGAATGCCAGGATTTGACCTTTACCTTCTAGTGGAGACAGAACTATGT 2661  
2341 CCTGTGATGAATTTATATCCCAACCTTATGAAACTGTGTGCAATCTCTCCAGAA 2400  
2662 CCTGTGATGAATTTATATCCCAACCTTATGAAACTGTGTGCAATCTCTCCAGAA 2721  
2401 GGCAGAGGAGGCTCACCGAGACTGCGGGTACAGCAGACAGCAAGAAATGTGAGAA 2460  
2722 GGCAGAGGAGGCTCACCGAGACTGCGGGTACAGCAGACAGCAAGAAATGTGAGAA 2781  
2461 GGCCTGGCTTTCGAGCAGTACCTGTTCTGATAAATGGAAGCCTGTGACCCCTCC 2520  
2782 GGCCTGGCTTTCGAGCAGTACCTGTTCTGATAAATGGAAGCCTGTGACCCCTCC 2841  
2521 TTCTGACAGAGCTCTGGTTACATTTCAAGAAAAATGTGATTTCCCTGCCCATTTGATTC 2580  
2842 TTCTGACAGAGCTCTGGTTACATTTCAAGAAAAATGTGATTTCCCTGCCCATTTGATTC 2901  
2581 AAGTTAAGCGATTTGCTAGTTGGGGTCTTGCAGTTCACTTGTGGAATGGAGTGAGA 2640  
2902 AAGTTAAGCGATTTGCTAGTTGGGGTCTTGCAGTTCACTTGTGGAATGGAGTGAGA 2961  
2641 ATTGATCCAAATGGCTTAAAGAAAACTTCAATGGAGGACGACCATGTCCCAAACTG 2700  
2962 ATTGATCCAAATGGCTTAAAGAAAACTTCAATGGAGGACGACCATGTCCCAAACTG 3021  
2701 GATCTCAAGAA-----TCAGGTACATGAGGAGTCCCATGTTACAGTGAGTGCAATCAG 2754  
3022 GATCTCAAGAAATCAGGCTCAGGTACATGAGGAGTCCCATGTTACAGTGAGTGCAATCAG 3081  
2755 TATTCCTGGGTTTACACACACTGTTCTCATGCAAAATCAACAAATGAGTGGGTCCTG 2814  
3082 TATTCCTGGGTTTACACACACTGTTCTCATGCAAAATCAACAAATGAGTGGGTCCTG 3141  
2815 CGCTGAGGAGGAAACAAATCTAGGAAAAATCAGATGTGGAATCTGCGGATGGTGA 2874  
3142 CGCTGAGGAGGAAACAAATCTAGGAAAAATCAGATGTGGAATCTGCGGATGGTGA 3201  
2875 GGTGAGCAGTGATGAGCAACCTGTGCAACCCAGGATGAAATTTCCCCCAGAAACCCAGTCC 2934  
3202 GGTGAGCAGTGATGAGCAACCTGTGCAACCCAGGATGAAATTTCCCCCAGAAACCCAGTCC 3261  
2935 TGTTCCTTATGTGTCCTCAATGATGTGTCATGTCGAGTGGGACCTTTGGAGCAAAATGC 2994  
3262 TGTTCCTTATGTGTCCTCAATGATGTGTCATGTCGAGTGGGACCTTTGGAGCAAAATGC 3321  
2995 CACAGTCAATGCAATCCCAACAAATGCGAGAGAACTCGGCACTCTGTATAGACATCA 3054  
3322 CCACAGTCAATGCAATCCCAACAAATGCGAGAGAACTCGGCACTCTGTATAGACATCA 3381  
3055 CTGAACCTCAAGGACTTGTGCTGAAGACTCACAGGTGCGCTTGCCTCTCTGAATGAAAT 3114

Db CTGAACCTCAAGGACTTGTGCTGAAGACTCACAGGTGCGCTTGCCTCTCTGAATGAAAT 3382  
QY TGCCTCCAGTTCAGTACATCTAAACAGAGTGGAGCACAATGAGTGGTGAAGAGCA 3174  
Db TGCCTCCAGTTCAGTACATCTAAACAGAGTGGAGCACAATGAGTGGTGAAGAGCA 3501  
QY CCTGTGCTCAAGCGCTCAGGACCCGCTCTAAGCTGTGTGTCAGTGTATGTCAGAGCA 3234  
Db CCTGTGCTCAAGCGCTCAGGACCCGCTCTAAGCTGTGTGTCAGTGTATGTCAGAGCA 3561  
QY GTACAGATGACCAATGTGAGCAGCATATTTGGAGAGCCCCCAGAGATGAGCATTTCC 3294  
Db GTACAGATGACCAATGTGAGCAGCATATTTGGAGAGCCCCCAGAGATGAGCATTTCC 3621  
QY TGCCTGTGTAATGCGTGTCAACTGTGACCTCTCAGGCTGGAGCGCTTGGACAGAGTGT 3354  
Db TGCCTGTGTAATGCGTGTCAACTGTGACCTCTCAGGCTGGAGCGCTTGGACAGAGTGT 3681  
QY TCACAGACCTGTGCGCATGAGAGTGCAGTGGAGCGCTTGCATTTATCATTTATGCCAAC 3414  
Db TCACAGACCTGTGCGCATGAGAGTGCAGTGGAGCGCTTGCATTTATCATTTATGCCAAC 3741  
QY CAAGGAGAGGAGCGGCGCATGCCACACAGAGCTTACCCAGGAGAAACCTGCCAGTGACC 3474  
Db CAAGGAGAGGAGCGGCGCATGCCACACAGAGCTTACCCAGGAGAAACCTGCCAGTGACC 3801  
QY CCTGTGCTACAGTGGGCTTGTGGCACTGTCTGCAATGTAATTTGAGAGGTGGAGACTGT 3534  
Db CCTGTGCTACAGTGGGCTTGTGGCACTGTCTGCAATGTAATTTGAGAGGTGGAGACTGT 3861  
QY GGGGAGAGGATTCAGATCCCGACGCTTCTGTCATGTCGACAGTGGTCAATATCTCAT 3594  
Db GGGGAGAGGATTCAGATCCCGACGCTTCTGTCATGTCGACAGTGGTCAATATCTCAT 3921  
QY GCAGCTGGAGCTGTGAGGATGCACTGTGTGAGAAATGCCCTTTTCAGGACAGCATCCTG 3654  
Db GCAGCTGGAGCTGTGAGGATGCACTGTGTGAGAAATGCCCTTTTCAGGACAGCATCCTG 3981  
QY AAGCAGCTGTGTCGTGTCCTTCCAGGAGACTGCCATTTACAGAAATGGTCCAGAGTGG 3714  
Db AAGCAGCTGTGTCGTGTCCTTCCAGGAGACTGCCATTTACAGAAATGGTCCAGAGTGG 4041  
QY AGCACAATGAAATTAACCTGCAATGTAAGAAAGCTTTGAGACTGTGGCCCGCAGTCT 3774  
Db AGCACAATGAAATTAACCTGCAATGTAAGAAAGCTTTGAGACTGTGGCCCGCAGTCT 4101  
QY AGATCAAGGACTTTTATAATTCAGTCTTTTGAACCAAGACAGCTGCCCCCAACAGGTT 3834  
Db AGATCAAGGACTTTTATAATTCAGTCTTTTGAACCAAGACAGCTGCCCCCAACAGGTT 4161  
QY CTAGAAACACGCTTGTACAGGAGGCAATGTTATCCTACATACATGGAAGCAAGTCTT 3894  
Db CTAGAAACACGCTTGTACAGGAGGCAATGTTATCCTACATACATGGAAGCAAGTCTT 4221  
QY TGGAAACAAATACGAACTGTATGTCGAGCTTCCAGTGGCTTAAATGTCAACAGGA 3954  
Db TGGAAACAAATACGAACTGTATGTCGAGCTTCCAGTGGCTTAAATGTCAACAGGA 4281  
QY GGTGCTCCCTCAGGCGCTTCTGTCGTCATTTGGCAGTGCATTTCCAGCTCCAGGAAA 4014  
Db GGTGCTCCCTCAGGCGCTTCTGTCGTCATTTGGCAGTGCATTTCCAGCTCCAGGAAA 4341  
QY CCTTCTCTTACTGTACAGAGGCTGAGTCTGTGTTTGTGAGAGGCTTATACAGAGATA 4074  
Db CCTTCTCTTACTGTACAGAGGCTGAGTCTGTGTTTGTGAGAGGCTTATACAGAGATA 4401  
QY ATGAATCAAAATGGTTTCTGGAATTTACTGCAATGAAGTACAGGCTCAGAGGATTAATA 4134  
Db ATGAATCAAAATGGTTTCTGGAATTTACTGCAATGAAGTACAGGCTCAGAGGATTAATA 4461  
QY GCTGATCTGAAACCTTTCTGGGAAAAACAGACCTGTGTAATTTCAAAAAATACATGATTT 4194

Db 4462 GCTGATGTGAAAAACCTTTCTGGGAAAAACAGACCTGTGAATTCAAAAATACATGATATT 4521  
QY 4195 TTTAAAGGATGGTCTCTTCAACCACTTGATCCAGATGGCGAGTAAATAATTTGGGTTTAT 4254  
Db 4522 TTTAAAGGATGGTCTCTTCAACCACTTGATCCAGATGGCGAGTAAATAATTTGGGTTTAT 4581  
QY 4255 GGCCTTTAGTGGCGCTTTCTCATCATGATTTTCCTTAATAATTTACTTCTACCTTGTT 4314  
Db 4582 GGCCTTTAGTGGCGCTTTCTCATCATGATTTTCCTTAATAATTTACTTCTACCTTGTT 4641  
QY 4315 TGCAGAGCCAAACACATCAAGACACACCTCCCAACAGAGGCTCTGACCTTAGCC 4374  
Db 4642 TGCAGAGCCAAACACATCAAGACACACCTCCCAACAGAGGCTCTGACCTTAGCC 4701  
QY 4375 TACGATGGAGACTTAGACATGTAA 4398  
Db 4702 TACGATGGAGACTTAGACATGTAA 4725

RESULT 2  
ID AAL43703 standard; cDNA; 3382 BP.  
XX  
AC AAL43703;  
XX  
DT 16-SEP-2002 (first entry)  
XX  
DE Human large protein 53-24 cDNA sequence.  
XX  
XX Human; gene; ss; large protein 53.24; embryonic development deformity;  
KW protein metabolism disturbance; tumour; immunologic system disturbance.  
XX  
XX Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 819..2273  
FT FT /\*tag= a  
FT FT /product= "Human large protein 53-24"  
XX  
PN CNI133246-A.  
XX  
PD 30-JAN-2002.  
XX  
PF 07-JUL-2000; 2000CN-00117029.  
XX  
PR 07-JUL-2000; 2000CN-00117029.  
XX  
PA (SHAN-) SHANGHAI BIODOOR GENE DEV CO LTD.  
XX  
PI Mao Y, Xie Y;  
XX  
DR WPI; 2002-305572/35.  
XX  
DR P-PSDB; AAO15279.  
XX  
PT New human large protein 53.24 and encoding polynucleotide, useful for  
PT treating embryonic development deformity, protein metabolism disturbance,  
PT tumor and immunologic system disturbance disease.  
XX  
PS Claim 6; Page 27-29 (Disclosure); 36pp; Chinese.  
XX  
CC The invention comprises the amino acid and coding sequence of the human  
CC large protein 53.24. The 53.24 DNA and protein sequences of the invention  
CC are useful for treating embryonic development deformity, protein  
CC metabolism disturbance, tumour and immunologic system disturbance. The  
CC present cDNA sequence encodes the human large protein 53.24  
XX  
SQ Sequence 3382 BP; 983 A; 705 C; 779 G; 915 T; 0 U; 0 Other;

Query Match 50.8%; Score 2235.8; DB 6; Length 3382;  
Best Local Similarity 99.6%; Fred. No. 0;  
Matches 2264; Conservative 0; Mismatches 2; Indels 7; Gaps 2;  
QY 2133 GCAGACAAACGGCAGCTCTCTCTGTGCAAGAATGCACAGTCCCTGATGCGAGAAGACTG 2192

1 GCAGACAAACGGCAGCTCTCTCTGTGCAAGAATGCACAGTCCCATGTGCGAAGACTG 60  
2193 CACCTTCTACTGTTGGTCCAAGTTTACGCCCTGCTCCACGAACTGTGAGCCACCAAAAG 2252  
61 CACCTTCTACTGTTGGTCCAAGTTTACGCCCTGCTCCACGAACTGTGAGCCACCAAAAG 120  
2253 TAGGCGGCGACAGCTCACAAGGAAAAAGCAGAAAAAGAGAGAAATGCCAGGATTTCTGACCT 2312  
121 TAGGCGGCGACAGCTCACAAGGAAAAAGCAGAAAAAGAGAGAAATGCCAGGATTTCTGACCT 180  
2313 TTACCTCTTAGTGGAGACAGAACTATGTCCTTGTGATGAATTTATATCCAACTTATGG 2372  
181 TTACCTCTTAGTGGAGACAGAACTATGTCCTTGTGATGAATTTATATCCAACTTATGG 240  
2373 AAACTGGTTCAGATTGCATTTCTTCCAGAAAGGAGAGGAGCTCACCCGAGACTGCGGGT 2432  
241 AAACTGGTTCAGATTGCATTTCTTCCAGAAAGGAGAGGAGCTCACCCGAGACTGCGGGT 300  
2433 ACAAGCAGACAGCAAGAATGTGGAGAAAGGCTTGGCTTTCGAGAGTAGCTGTTCTGA 2492  
301 ACAAGCAGACAGCAAGAATGTGGAGAAAGGCTTGGCTTTCGAGAGTAGCTGTTCTGA 360  
2493 TAAAAATGCAGACCTGTTGACCCCTCTCTCTGAGAGAGCTCTGTTTACATTCAAGAAA 2552  
361 TAAAAATGCAGACCTGTTGACCCCTCTCTCTGAGAGAGCTCTGTTTACATTCAAGAAA 420  
2553 ATGTGTCATTCCTGCCCATTTGATTGCAAGTTAAGCAATGGTCTAGTTGGGGGCTTTG 2612  
421 ATGTGTCATTCCTGCCCATTTGATTGCAAGTTAAGCAATGGTCTAGTTGGGGGCTTTG 480  
2613 CAGTTTCATCTTGTGGAATTTGGAGTGAGAAATCGATCCAAATGGCT-AAAAAGAAAACCTT 2671  
481 CAGTTTCATCTTGTGGAATTTGGAGTGAGAAATCGATCCAAATGGCTTAAAAAGAAAACCTT 540  
2672 ACAATGGAGGACGACCATGTCCCAACTGGATCTCAAGAA-----TCAGTACATGAGG 2725  
541 ACAATGGAGGACGACCATGTCCCAACTGGATCTCAAGAACTCAGGCTCAGGTACATGAGG 600  
2726 CAGTCCCATGTTACAGTGAATGCAATCAGTATCTCTGGTTGTAGAACACTGCTTCTCAT 2785  
601 CAGTCCCATGTTACAGTGAATGCAATCAGTATCTCTGGTTGTAGAACACTGCTTCTCAT 660  
2786 GCAAAATCAACAATGAGCTGAGGTCCTCTGCGCTGTGGAGGAGGAACAACAATCTAGAAAA 2845  
661 GCAAAATCAACAATGAGCTGAGGTCCTCTGCGCTGTGGAGGAGGAACAACAATCTAGAAAA 720  
2846 TCAGATGTGTGAATGAGTGGGATGTTGAAGTGGAGCAGTGAATGAGCAACTGTGCAACC 2905  
721 TCAGATGTGTGAATGAGTGGGATGTTGAAGTGGAGCAGTGAATGAGCAACTGTGCAACC 780  
2906 AGGATGAATTCCTCCAGAAACCCAGTCTCTCTCTTATGTGTCCTCCCAATGAGTGTGCA 2965  
781 AGGATGAATTCCTCCAGAAACCCAGTCTCTCTCTTATGTGTCCTCCCAATGAGTGTGCA 840  
2966 TGTCTGAGTGGGACTTTGGAGCAAAATGCCACAGTCAATGCCACCAACAATGCGAGA 3025  
841 TGTCTGAGTGGGACTTTGGAGCAAAATGCCACAGTCAATGCCACCAACAATGCGAGA 900  
3026 GAAGAACTCGCCACCTGCTTAAGACCATCACTGAATCAAGGACTTGTCTGAAGACTCAC 3085  
901 GAAGAACTCGCCACCTGCTTAAGACCATCACTGAATCAAGGACTTGTCTGAAGACTCAC 960  
3086 AGGTGAGCTTGGCTCTCTGAAATGAAATTTGTTCCAGTTCCAGTACAAATCTTAACAGAGT 3145  
961 AGGTGAGCTTGGCTCTCTGAAATGAAATTTGTTCCAGTTCCAGTACAAATCTTAACAGAGT 1020  
3146 GCAGACATGCCAGCTGAGTGAAGACGACCTGTGGTCAAGGCTCAGGACCCGCTGCG 3205  
1021 GCAGACATGCCAGCTGAGTGAAGACGACCTGTGGTCAAGGCTCAGGACCCGCTGCG 1080  
3206 TAAGCTGTGTGTCAGTGAATGGAAGCCAGTCAAGATGCAATGTGAGCAGCATAAAT 3265





CC The present invention relates to novel human NOV proteins and their  
CC coding sequences (ACC72075-ACC72181 and AB858469). The NOV  
CC proteins are useful in manufacturing a medicament for treating a syndrome  
CC associated with a human disease. The NOV proteins and coding sequences  
CC may be used to diagnose, treat or prevent metabolic disorders such as  
CC diabetes or obesity, infections, cachexia, cancer, neurodegenerative  
CC disorders such as Alzheimer's disease or Parkinson's disease, immune  
CC disorders, haematopoietic disorders and various dyslipidaemias  
XX  
SQ Sequence 4810 BP; 1351 A; 1104 C; 1317 G; 1038 T; 0 U; 0 Other;

Query Match 28.0%; Score 1232.8; DB 7; Length 4810;  
Best Local Similarity 57.7%; Pred. No. 0;  
Matches 2598; Conservative 0; Mismatches 1782; Indels 126; Gaps 17;  
QY 10 TGCATTTCAGAGCTGAACCGAATCTGGTGTGCAATGAAATATGCAACACTTTGGCCCTT 69  
DB 288 TGCAATCAGAAAGACAAGACATCTCTGGGAGATATCATCTGTGAGTACTTTGAGCCC 347  
QY 70 CAGCTCTCTACAGAACAGGCTTGCTCATCTCTTGTCCCGGATTTGTGATATCTGAG 129  
DB 348 AAGCCTCTCTGAGACAGGCTTGCTCATCTCTTGTCCCGCAAGATTGCATCGTGTCTGAA 407  
QY 130 TTCTTACATGCTCCCACTGTAGCAAGGATGTGGGAAGAAATTCAGCATACAACTCCG 189  
DB 408 TTTTCTGCTGGTCCGAATGCTCAAGACCTGGCGAGCGGCTCCAGCACCGGACGCGT 467  
QY 190 GCGGTCAATAGCTCCCTCTCTTTTGGTGTGTTGCAATGTCCAAATCTGACTGATCAAGA 249  
DB 468 CATGTGTGGCGCCCGCCAGTTCGGAGGCTCTGGCTGTCCAAACCTGACGGAGTTCACG 527  
QY 250 GCCTGTGATCTCCCATTTCTGCTCTCTTGGGAGAGGAATATACATTTAGCTTAAAG 309  
DB 528 GTGTG-----CNAATCCAGTCCATGCGAGCGGAGGCTCAGGTACAGCCCTGCAAT 578  
QY 310 GTTGGACATGGAGTAAATGAGACTGCTCATCTTAAAGAAATTAATCCAAAGCGGAAGA 369  
DB 579 GTGGGCGCTGGACACCTGCTCAATGSCCCACTCCCGACAGTAAGACAAGCAAGGAGA 638  
QY 370 ACTGTTCTGANTTTAA-----CTCTGATTCAAATGAG 402  
DB 639 CGCGGGAAGAAATAAGAACCGGAAAGGACCGCAGCAAGAGGTAAGGATCCAGAGCC 698  
QY 403 CGAGTCACTTTAAACATCAAAGTTACA-----AAGCACATCATCATTCGAAGTCT 453  
DB 699 CGGAGCTTATTAAGAAAGAGAAACAGAAAAGGACAGACAGACAGACAAATAT 758  
QY 454 TGGCAATAGAGATAGTTATCAAAACCGGAGGTTTCGTGTACAGAAAGTGTAGACAA 513  
DB 759 TGGGACATCCAGATTGGATATCAGACAGAGAGGTTATGTGCATTAAACAAGCGGGAAA 818  
QY 514 AATGCTATGTTAGCCCTTGCCTTCAAGATTCCTCCATTGACTGTTCAGTCCCTGCATC 573  
DB 819 GCTGCTGATTTAAGCTTTTGCAGCAGAGAGAGCTTCCAAATGACCTTCCAGTCTCTGTG 878  
QY 574 ATGCCCAAAGACTGTGAACCTCCAGTGTCTCTTGGAGCCCTGCTCCAAAGACATGC 633  
DB 879 ATCAACAAAGAGTCCAGGTTTCGAGTGTGTAGAGTGGAGCCCTGCTCAAAAACATGC 938  
QY 634 CGTTCAGGAGTCTCTTGCAGGATTTAGGACAGGAGCCGGAACGTGAGACATGCT 593  
DB 939 CATGACATGTTGCTCCCTCGAGGACCTCGTGTAGGACACGAACCATCAGGAGTTCC 998  
QY 694 ATTGGAGGTGGAAGAGTGTCTCTGAACTTCTTGAGAAAGAGCCCTGCATTTGTGAGGA 753  
DB 999 ATTGGCAGTGAAGAGAGTGTCCAGAAATTTGAAGAAAAGAAACCTGTTGTCTCAAGGA 1058  
QY 754 GAATCTTCAGACATGCTCCAGGATTTCTGAGAACTTCTGAAATGGAAGAAATGCCAA 813  
DB 1059 GATGGAGTTGTCCCTGTGCAAGATATGCTGAGAACTACAGAGTGGAGTGAAGTGCCT 1118  
QY 814 GTCTCTCTCTCTCCGAGCAGAGATCCCACTGGCATGTGACGGGACCCGCTGTGTGGC 873

DB 1119 GTGGACCTTTGCTCAGTCAGCAGACAGAGGCGCGCAACACAGACGCGCCTCTGTGGA 1178  
QY 874 GGTGGATCCAGACCCGGAGGTGTACTGTGCCAG-----AGGTACACGCA 921  
DB 1179 GGGGCAATCCAGACCCGAGAGGTGTACTGTGGTGAGGCCAAACGAAAACCTCTCTCAAA 1238  
QY 922 GCTGCGGCACTGAGGCGCAAGAAAGTCTCTAGACCTGTGAAAGGCAATTTATGTGTGGGA 981  
DB 1239 TTAAGTACCCACAAGAACAAAGAGCCTCAAGCAATGGACTTAAATATATGCACTGGA 1298  
QY 982 CCGGCCCGTGGCCCTCTCAGCTCTGCAATATCCCTTGTCTTACGAGTGCATATCT 1041  
DB 1299 CCTATCCCTAATATCTACACAGCTGTGCCAATCTCTTGTCCAACTGAATGTGAAGTTCA 1358  
QY 1042 TCCTGCTCAGCCTGGGCGCTGTGCATCATGAAACCTGTATGAACCTCAGGGGAAAAA 1101  
DB 1359 CTTGGTTCAGCTTGGGACCTTGTACTTATGAAACCTGTATGATCAGCAAGGAAAAA 1418  
QY 1102 GGATTTAGAACGAGCAGCG-----CATGTCTCATGGAATCTACAGGCGCTGCA 1152  
DB 1419 GGCTTCAAACCTGAGAAAGCGCGCAATTACCAATGAGCCCACTGGAGCTCTGGGGTAACC 1478  
QY 1153 GGGCAATCCCTCATTTGGTGGAGTCTGTCTTGTGAGGATCCAAATGTGTACTCCGATGG 1212  
DB 1479 GGAACCTGCCCTCACTTACTGGAAGCCATTCCTCTGTGAGAGCTGCTGTATTGACTGG 1538  
QY 1213 ---CTGCGCATCAGAAAGGATCTGTTTCCCTGATCATGAA---AATGTGGCCTGGGACAT 1266  
DB 1539 AAAGCGGTGAGACTGGGAGACTGGGAGCAGATAACGGAAGAGAGTGTGTCTCCAGGACG 1598  
QY 1267 CGTATTCGAGGCGCTGCGCAGAAATGACCGGAGAAAGTATCAGGGAGTCTTTGC 1356  
DB 1599 CAAGTTCAAGAGGTTGTGTGCATCAACAGTATGGAAGAAAGTTGACAGACACTGTGC 1658  
QY 1327 CAGTTCCTCCCTCTCTGAGAGGAAGTCTTGTGAAATTCCTCCGAATGGACTGTGTG 1386  
DB 1659 AGAGATGCCATCTTCCCATCCCTGTGGCCTGTGATGCCCATCCCGAAAGACTGTGTG 1718  
QY 1387 CTGAGCAGTGGAGGAGTGTCTCTGTCTCCAGCTCTGTTCCTTCAATTAATACTCAGAT 1446  
DB 1719 CTCAGCAGATGGTCTAGCTGTCTCTCTGTCTCAGACCTGCTCAGGGAAACACAGAA 1778  
QY 1447 GGGAAACAGACAGCTCAAGAACTATCTGGC---ACTGGCTGGGGAAGGTGGAAGCACA 1503  
DB 1779 GGGAAACAGATACAGACAGATCCATCTGGCCTATGCGGTGAAGAGGTGGAATTCGC 1838  
QY 1504 TGTCCCTCTAGTACGCTCTCCAAAGAGCATGTTGTGTGTAATGACCATTCCTGTATGAG 1563  
DB 1839 TGTCCAAATAGCAGTGTCTTGAAGAGAGTACGAAGCTGTAAATGAGCATCTCTGACAGT 1898  
QY 1564 CTTCACTGGGAGACATCGCCTTGGGCGCTTGTCTGAGGACACATTTGTAACCTGCGCTT 1623  
DB 1899 TACACTGGCAAACTGGTCCCTGGGCGCAGTGCATGAGGACACTCAGTATCGTCTTC 1958  
QY 1624 AATGCAACATTTGGCTGGAATGGAAGACCACTGTGTGTGTGAGCATTGAGCTCGGAGA 1683  
DB 1959 AACCAACTAGCACTTGGAAATGGGAGGCTCTCTGCTCTGTGGCATGCAAGAAAA 2018  
QY 1684 GTCTTCTGTGTCAAGAGTCACTGCGGCAAGTAAATGACCAAAAGATGTCCAGATTTACT 1743  
DB 2019 GTCATCTGTGTGAGTCAATGTGGGCGAAGTGGGACCCAAATAATGTCTCTGAAGCCTT 2078  
QY 1744 CGACTGAAACTGTGCGCCCTGTTTCTCCCATGCAAAAAGACTGTATTGTGACTGCT 1803  
DB 2079 CGACTGAAACTGTGAAAGCCTTGTCTGCTTCTTGTGAAGAGACTGTATTGTGACCCCA 2138  
QY 1804 TTCAGTGTGAGTGCACCTGCGCAAGATGTGCCAAGCAGGAATGCCACAGT---AAAA 1860  
DB 2139 TATAGTGTGAGTGCATCATGCCCTCTCTGCTGTGAAGAGGGGACTCCAGTATCAGGAAG 2198  
QY 1861 CAGTCTCATACAGAAATCATCATCAAGAGACAGCCCAATGGAGGCCAGGAATGCCACAGAT 1920  
DB 2199 CAGTCTAGGCATCGGGTCATCATTCAGCTGCCAGCAACCGGGGCGGAGACTGCACAGAT 2258

Qy	2992	TGCCACAGTCTGCAGTCCCAACAAT---GCAGAGAAAGAACTCGCCACCTGCTTAAGA	3048
Db	3339	TGTGTTTTGCTTTGCCAATCAAGACAGTTTCGGCAAGAGTTCAGCTGATCCATCAGACAA	3398
Qy	3049	CCATCACTGAACTCAAGGACTTGTGTGTAAGACTCAAGGTGCAGCGCTTCCCTCCCTGAAT	3108
Db	3399	CCAGCTGATGAAGAGATCTTCGCCCTAATGCTGTTGAGNAAGAACTCCTGTAACTTGAAC	3458
Qy	3109	GAAAAATTGCTTCCAGTTCCAGTACAATCTTAACAGAGTGGAGCAATGCCAGCTGAGTGAA	3168
Db	3459	AAAACTGCTTACCACATGATTAATATGTAACAGACTGGAGTAGTACATGTCACTGTAGTGAG	3518
Qy	3169	AAGCACCCTGTGTCAAGGCGTCAGACCGCTCTCTAAGCTGTGTGTCAGTGTGTC	3228
Db	3519	AAGCGAGTTTGTGGAATGGAATAAANAAGAGATGTGGAATTTGTGTTGGAAGTGAATGCG	3578
Qy	3229	AAGCCAGTCAGCATGGAACCAATGTGAGCAGCAATAATTTGGAGAAGCGCCCAAGAGAAATGAGC	3288
Db	3579	AAGTCAGTTGACCTGAAATATTGTGAAGCGCTTGGCTTGGAGAAGAACTGGCAGATGAAC	3638
Qy	3289	ATTCCTCGCTTGTGGAATGCGTGTCTCACTGTTCAGCTCTCAGGGTGGACGGCTTGGACA	3348
Db	3639	ACGTCTGCAATGTGGAATGCCCTGTGAACTGTCAAGCTTCTGATTGGTCTCTCTGGTCA	3698
Qy	3349	GAGTGTTCACAGACCTGTGGCCATGGAGGTTCGAATGAGCGCGACTCGAATTTATCATTTATG	3408
Db	3699	GAATGTTCTCAAAACATGTGSCCTCA CAGGAAATAATGATCCGAGACGAACAGTGA CCCCAG	3758
Qy	3409	CGAACCCAGGAGAAGACGGCCATCGCCCA CAGAGCTTACCCAGAGAGAAAACCTGCCCCA	3468
Db	3759	CCCTTTCAAGGTGATGGAGA CCAATCCCTTCCCTGATGGAC CAGTCCAAACCCCTGCCCCA	3818
Qy	3469	GTGACCCCTCTACAGCTGGGTCTTGGCAACTGTCTGCATGTAAATTTGGAGGGTGGGA	3528
Db	3819	GTGAGCCTTGTTATCGTGGGCAATATGGCCAGTGTCTCCATGCCAGTGCAGAGAGGCC	3878
Qy	3529	GACTGTGGGAAGGAGTTCCAGATCCGACGCTTTCTGTGATGGTCCACAGTGGTTCAATA	3588
Db	3879	CAGTGTGGAGAAGGAC CAGAA CAAAGAA CATTCTTGTGTAGTAAGTGAATGGGT CAGCT	3938
Qy	3589	TCTCATGCAGCTGGACGTGTCCAGAGTGCACGTGTGGAGAAATGCCCCCTT-----	3638
Db	3939	GATGATTTCCAGAAAGTGTGGATAGGAAATCTGTGCTGACATGGAATCAATATAGAT	3998
Qy	3639	-----TCAGACAGCATCCTGAAGCAGCTGTGTGTGTGCTGCTTGCCTCCAGGAGACTGCCAT	3693
Db	3999	GGTAATAAAAAATATGTTCTGGAGGAATCCTGCAGCCAGCTTGGCCAGGTGACTGTAT	4058
Qy	3694	TTAACAGAAATGTTCAAGTGGAGCACAATGTGAATTAACCTGCATGATGGAAGAACCTTT	3753
Db	4059	TTGAAAGGACTGTFCTTCCCTGGAGCCGTGTGCAGCTGACCTGTGTGAATGGTGAAGGATCTA	4118
Qy	3754	GAGACTGTGGGCGCCAGTCTAGATCAAGACCTTTTATAATTCAGTCTTTTCAGAACCCAA	3813
Db	4119	GGCTTTGGTGGAAATACAGTCCAGACCGGTGATTTATACAGAACTAGAGAAATCAG	4178
Qy	3814	GACAGTGCCTCCCAACAGGTTCTAGAAAACGCGCTTGTACAGGAGGCAAAATGTTATCAC	3873
Db	4179	CATCTGTGCCAGAGCAGATGTTAGAAAACAAAATCATGTTATGATGGACAGTGTCTATGAA	4238
Qy	3874	TACACATGAAGAACGAGTCTTTGGACAATAACGAAACGAACTGTATGTTGCCAGCGTTCA	3933
Db	4239	TATTAATGGAATGGCCAGTCTCTTGGAAAGGGCTCTTCCGAAACAGTGTGGTGTCAAAGGTCA	4298
Qy	3934	GATGGCGTTAATGTACAGGAGGCTGCTCCCTCAGGCCCGCTCTGCTGCTGCCATTCGGCAG	3993
Db	4299	GATGGTATAAAATGATAACAGGGGGCTGCTTTGGTGTGATGAGCCAGCTGATGCCAGAGTCT	4358
Qy	3994	TGCATTCAGCCTGCAGAAAACTTTCTCTCTACTGTACACAGGGTGGAGTCTGTGTGGTGT	4053
Db	4359	TGTAAACCCACCGTGTAGTCAACCCCACTGTAATGTGTAGCGAGACAAAAATGTGCCATTTGT	4418
Qy	4054	GAGAAGGGGTATACAGAGATAATGAAATCAAAATGGTTTCTCTGATTTACTGCA-----	4105



1313 CATGACATGGTGTCCCTCGAGGCACTCGTGTGAAGACAGCAACCAATCAGGCGAGTTTCCC 1372  
694 ATTGGAGTGGAAAGAGTGTCTCTGAACCTCTTGTGAAAGAGCGCTGCATTTGTGAAGGA 753  
1373 ATTGGCAGTGAAGAGAGTGTCCAGATTTGAAGAAAAGAACCCCTGTTTGTCTCAAGGA 1432  
754 GAACCTCTCGACAGCAATGCCAGGTATCTCTGAGNACTTCTGAATGGAAGAAATGCCAA 813  
1433 GATGGAGTTGTCCCTGTGCCAGTATGGCTGGAGAACTACAGATGGACTGAGTGGCGGT 1492  
814 GTCTCTCTCTCTCTCTCGAGCAGCAGGATCCCACTGGCATGTGACGGGACCCGTGTGTGCG 873  
1493 GTGGACCTTTGTCTCAGTCAGCAGGACAGAGCGCGCAACCAAGCGCCCTCTGTGGA 1552  
874 GTGGGATCCAGACCCGGGAGGTGTACTGTGCCAG-----AGGTCACAGCA 921  
1553 GGGGGCATCCAGACCCGAGAGGTGTACTGCTGTCAGGCCCAACGAAAACCTCTCTCACAA 1612  
922 GCTGCCGCACTGAGGCGCCCAAGAAAGTCTCTAGACCTGTGGAAGAGGCAATTATGTGGGA 981  
1613 TTAAGTACCAACAGAACAAAGAGGCTCAAAGCCATGGACTTAAATTTATGCACTGGA 1672  
982 CCGGCCCTGTCCCTCTCAGCTCTGCAATATCCCTGTCTTACGGACTGCAATATCT 1041  
1673 CCTATCCCTAATACTACACAGCTGTGCCACATCTCCTGTCCAACTGAATGTGAAGTTCA 1732  
1042 TCTGTGTGAGCTGGGCGCTGTGCATCCATGAATACTGTCAACCTCAGGCGGAAAGAAA 1101  
1733 CTTGTGTGAGCTGGGAGCTTGTACTTATGAAGAACTGTGAATGANTCAGCAAGGGAAAGAA 1792  
1102 GGATTTAGAACAGGCGAGCG-----CCATGTCTCTCATGGAATCTACAGGCGCTGCA 1152  
1793 GGCTTCAAACCTGAGGAAGCGCGCATTAACCAATGAGCCCACTGGAGGCTCTGGGGTAACC 1852  
1153 GGGCATTTGCCCTCATTTGTGTGAGTGTCTTGTGTGAGGATCCAAATGTGTACCGATGG 1212  
1853 GGAATCTGCCCTCACTTACTGTGAAGCCATTCCTGTGTGAAGAGCTGCTGTATGACTGG 1912  
1213 ---CTGGCATCAGAAAGGATCTGTTTCCCTGTATCATGGA---AATGTGGCTGGGACAT 1266  
1913 AAGCGGTGAGACTGGGACCTGGAGCCAGATACAGGAAGGAGTGTGTCTCAGGACAG 1972  
1267 GCTATTTGGAAGCGCTGTGCAGAAATGACCGGGAGAGATGTATCAGGAGTCTTTGC 1326  
1973 CAAGTTCAAGAGTGTGTGTGCATCAACAGTGTATGGAAGAAAGTTGACAGACAGCTGTGC 2032  
1327 CCAGTTTCCCTCTCTCGAGAGGAGTCTTGTGAAATTCCTGCGCAATGGAAGTGTGTG 1386  
2033 AGAGATGCCATCTTCCCATCCCTGTGGCTGTGATGCCCAATGCCGAAGAGCTGTGTG 2092  
1387 CTGAGCGAGTGGACGAGTGTGTCTCTGTTCCAGTCTCTGTTCAAAATAAAACTTCAGAT 1446  
2093 CTCAGCACATGTGTACGTGTCTCTCTGCTCACACACTGCTCAGGGAAGAACGACAGAA 2152  
1447 GGGAAACAGACAGTCAAGAACTATCTGGC---ACTGGCTGGGGAAGTGTGAAGCCA 1503  
2153 GGGAAACAGATACGAGCAGATCCATCTTGGCTATGCGGGTGAAGAGGTGGAATTCGC 2212  
1504 TGTCCCTCTAGTCAGGCTCTCCAAGAGCATCGTTGTGTATGACCATTCCTGTATGCGAG 1563  
2213 TGTCCAAATAGCAGTGTCTTGAAGAGTACGAAGCTGTATGAGCATCTTGCACAGTG 2272  
1564 CTTTCTGAGGACATTCGCTTGGGGCCCTTGTCTGTGAGGACACATTCGTAACCTGCGCTT 1623  
2273 TACCACTGTGCAAACTGGTCTTGGGGCCAGTGCATTTGAGGACACCTCAGTATCGTCTTC 2332  
1624 AATGCAACCATTTGGTGAATGAGAGCCAGCTGTGTGTGAGGCAATTCAGACTCGGAGA 1683  
2333 AACACAACTACGACTTGAATGGGAGGCGCTCTGCTGTGCGCATTCGACAGCAAGAAA 2392  
1684 GTCTTCTGTGTCAAGAGTCACTGGGCAAGTAATGACCAAGATGTCCAGATTTCTACT 1743  
2393 GTCATCTGTGTGCGAGTCAATGTGGGCCAAGTGGGACCCAAAGAAATGTCTTGAAGCCTT 2452

1744 CGACCTGAACCTGTGGCCCTCTGTTTTCTCCATGCAAAAAAGACTGTATTTGACTGTCT 1803  
2453 CGACCTGAACCTGTAGGCGCTTGTGCTTCTTGTAAAGAGGACTGTATTTGACCCCA 2512  
1804 TTCAGTGTAGTGGACACCTCTGCCCAAGATGTGCCAAGCAGGAAATGCCACAGT---AAAA 1860  
2513 TATAGTGTAGTGGACATCATGCCCTCTCTGCTGTAAAGAGGGGACTCCAGTATCAAGAA 2572  
1861 CAGTCTCGATACAGAAATCATCATCCAAAGAGCAGCAATGGAGGCCAGGAAATCCCCAGAT 1920  
2573 CAGTCTAGGATCGGTTCATCATTCAGTCTCCAGCCAAAGGGGCGGAGACTGCAAGAT 2632  
1921 ACCTTATATCAGGAGAGAGTGTGAAGATTTCTTGTGTCTGTATATCCGTGTGAAG 1980  
2633 CCCCTCTATGAAGAGAGGCGCTGTGAGGCACCTCAAGCGTGCACAAAGCTACAGGTGGAAG 2692  
1981 CCACAGAAATGGAGCCCTTTCATCTTAGTGCAGAGTCTGTCTGCGAGGGAATAACGGGC 2040  
2693 ACTCAAAATGGCGAGATGCCAATTAGTCCCTTGGAGCGTGCAACAAGACAGCCCTGGA 2752  
2041 AGCAGTGAAGCTTGTGAAAGGGTTTACAAACAAGAGCTGTCTCATGCATCTCTGTATGAC 2100  
2753 GCACAGGAAGGCTGTGGCGCTGGCGCAGCAGGCAAGAGCCATTACTTGTGCAAGCAAGAT 2812  
2101 AACCGTTCAGCAGAAATGATGGAATGCCCTCAAGCAGACAAACGGCATGCTCTCTTTGTG 2160  
2813 GGAGGACAGGCTGGAATCCATGAGTGCCTACAGTATGACAGGCCCTGTGCCAGCCCTTACC 2872  
2161 CAAGAATGCACAGTCCCATGTTCGAGAAGACTGCACCTTCACCTTGTGTCCTCAAGTTTACG 2220  
2873 CAGGCTTCCAGATCCCTGCCAGGATGACTGTCAATTCAGCAGCTGTGTCCTCAAGTTTCT 2932  
2221 CCCTGCTCCAGAACTGTGTGAAGCCACAAAGATGAGCGGGGACAGCTCAGAGGAAAGC 2280  
2933 TCATGCAATGAGACTGTGTGTCAGTTAGACCAGAAAGCGCACTCTTGTGTGAAAAAGT 2992  
2281 AGAAAGAGCAGAAATCCAGAGATCTGACCTTTACCTTTACCTCTAGTGAGACAGAACTATGT 2340  
2993 AAAAGAGGAAATGTAAATTTCCATTTGTATCCCTGATTCCTGATTCAGACTCAGTATTGT 3052  
2341 CTTGTGATGAATTTATATCCCAACTTATGGAACCTGTGTCAAGTTGCAATTTCTTCCAGAA 2400  
3053 CTTGTGACAAATATATATGCAACCTGTGGGAACTGTGTCAAGCTGTATTTTACCAGAG 3112  
2401 GGCAAGGAGGCGCTCACCGAGACTGCGGTACAGCAGACAGCAAGAAATGCTGTGAGAA 2460  
3113 GGAAGTGGAAAGTGTGTGGGAATGAAGATCAAGGAGACATCAAGGAATCCGAGCAA 3172  
2461 GGCTTCCGCTTTTCAGCAGTAGCTCTTTCTGATAAAATGGAAGACCTTGTGACCCCTCC 2520  
3173 GGATATCGTTTACCAAGCAATGGCATCTACGATCAAAATGGCAGGCTTGTGAAACATCT 3232  
2521 TTCTGAGCAGCTCTGGTTACATTCAGAAATGTGTCAATTCCTGCCCCCATTTGATTCG 2580  
3233 AGATGTAAACAGCCATGGTTACATTTGAGGAGGCGCTGCATCATCCCTGCCCCCTCAGACTGC 3292  
2581 AAGTTAAGCGAATTTGTAGTTGGGGCTCTTGCAGTTCATCTTGTGGAATTTGAGTGTGAGA 2640  
3293 AAGCTCAGTGTAGTGTCCAACTGTCGCGCTGCAGCAAGTCTCTGTGGGAGTGGTGTGAAG 3352  
2641 ATTTCGATCCAAATGGCTTAAAGAAACCTTACATGAGGAGGAGCAGCATGTCTCCAAACTG 2700  
3353 GTTCTGTTTAAATGGCTGTGAAAAACCATATAATGAGGAGGCGCTTGTGCCCAAAGT 3412  
2701 GATC-----TCAAGAAATCAGGTACATGAGGAGTCCCATGTTTACAGTGTGCTCAATCAG 2754  
3413 GACCATGTCAACCGGACAGGTTGTATGAGTGTGTCCTCCATGCCACAGTGTGCAACAG 3472  
2755 TATTCCTGGTGTAGAACACTGGTCTTTCATGCAAAATCAACAAATGAGCTGAGGTCCCTG 2814  
3473 TACCTATGGTTCACAGAGCCCTGGAGCATCTGCAAGGTGACCTTTTGTGAATATGCGGAG 3532

```
QY 2815 CGCTGTGGAGGAGGAAACAATCTAGGAAATCAGATGTGT---GAATCTCGGATGGT 2871
Db 3533 AACTGTGGAGGGCGTGCACAAACCGGAAAGTGAGATGCATGCAGATAACAGCAGATGGC 3592
QY 2872 GAAGGTGGAGCAGTGGATAGCAACCTCTGCACACAGGATGAAATCCCCAGAAACCCAG 2931
Db 3593 CCTTCTGAAACATGTAGAGATTACTCTGTGACCCAGAGAGATGCCCTTGGGCTCTAGA 3552
QY 2932 TCCCTGTTCTCTATGTGTCCCAATGAGTGTCTATGTCTGAGTGGGACATTTTGGAGCAA 2991
Db 3653 GTGTGCAAAATACCATGCCCTGAGGACTGTGTGATATCTGAATGGGGTCCATGGACCAA 3712
QY 2992 TGCCACAGTATCGGATCCCAACAAAT---GCAGAGAGAACTGCCACCTCTCAAGA 3048
Db 3713 TGTGTTTTCCTTTCGAAATCAAGCAGTTTCGGCAAGGTCAGCTGATCCCATCAGACAA 3772
QY 3049 CCATCACTGAATCAAGGACTTGTGCTGAAGACTCACAGGTGCAGCCTTGCCCTCCCTGAAT 3108
Db 3773 CCAGCTGATGAAGGAGATCTTGCCCTAAATGCTGTTGAGAAAGAACCTGTAACTGAAAC 3832
QY 3109 GAAATTTGCTTCCAGTTCAGTACAACTTAACAGAGTGGACACATGCCAGCTGAGTGAA 3168
Db 3833 AAAAACTGCTACCACTATGATTAAATGTAAACAGACTGGAGTACATGTCACTGAGTGAG 3892
QY 3169 AACGCACCTGTGTCTAAGGCTCAGAACCCGCTGCTAAGCTGTGTGTCAGTGTATGGC 3228
Db 3893 AAGCAGTTTCTGGAATGGAATTAACAAAGGATGTGGATTGTTCGAAGTATGGC 3952
QY 3229 AAGCAGTCAAGATGGACCAATGTGAGCAGCAATTAATTTGGAGAGCCCCAGAGATGAGC 3288
Db 3953 AAGTCAGTTGACCTGAAATATTGTGAAGCGCTTGGCTTGGAGAGAACTGGCAGATGAAC 4012
QY 3289 ATTCCTCTGTTGGTGAATGGTGTCAACTGTCAAGTCTCAGGCTCAGGAGTGGGACCA 3348
Db 4013 ACGTCTGCAATGGTGAATGCCCTGTGAATCTGAGCTTCTGATGCTCTCCTTGTGCTCA 4072
QY 3349 GAGTGTTCACAGACCTGTGGCCATGGAGGTGCAATGAGCCGAGCTCGAATTTATCATATG 3408
Db 4073 GAATGTTCTCAACATGTGGCTCACAGGAAATATGATCCGAGCAACAGTGACCCAG 4132
QY 3409 CCAACCCAGAGAGAGAGGAGGCGGATGCCCAACAGACTTACCAGGAGAAACCTGCCCA 3468
Db 4133 CCTTTCAAGGTGATGGAGAGACCATGCCCCCTTCCCTGATGGACCATCCAAACCCCTGCCCA 4192
QY 3469 GTGACCCCTCTCAGCTGCGGTCTTGGCAACTGGTCTGCATGTAAATTTGGAGGGTGGGA 3528
Db 4193 GTGAAGCCTTGTATCGTGGCAATATGGCCAGTGGTCTCCATGCCAAGTGCAGGAGGCC 4252
QY 3529 GACTGTGGGGAAGGAGTTTCAGATCGCGAGCCTTCTCTGATGTCATGTCACAGTGGTTCAATA 3588
Db 4253 CAGTGTGGAGAGGGACACAGAACAGGAACATTTCTTGTGTAGTAAGTGTATGGGTCAAGT 4312
QY 3589 TCTCATGCACTGGACGTGTGCGAGTATGCACTGTGTGGAGAAATGCCCTT----- 3638
Db 4313 GATGATTTACGAAAGTGTGTGATGAGGAATCTGTGCTGACATTTGAATCATATAGAT 4372
QY 3639 -----TCAGGACAGCATCCTGAAGCAGCTGTGTTCTGTGCTTCCAGGAGACTGCCAT 3693
Db 4373 GGTAAATAAATAATGTTTCTGGAGGAATCCTGTGACCCAGCAGCCTTCCAGGAGTACTGTTAT 4432
QY 3694 TTAAACAGAAATGTCAGAGTGCAGACATGTGCAATTAACCTGATTTGATGGAAGAGCTTT 3753
Db 4433 TTGAAGGACTGTTCTTCTGGAGCCTGTGTGAGTGTGCTGAGTGTGATGAGGATCTA 4492
QY 3754 GAGACTGTGGCCCGCAGTCTAGATCAAGGACTTTTATTAATTCAGTCTTTTGGAGAACCAA 3813
Db 4493 GGTCTTGGTGAATATACAGGTCCAGATCCAGACCGGTGATTATACAAAGAACTAGAGAATCAG 4552
QY 3814 GACAGCTGCCCCACAGGTTCTAGAAACACGCCCTTTGTACAGGAGCAATGTTATCAC 3873
Db 4553 CATCTGTCCACAGACAGATGTTAGAAACAAATATCATGTTATGATGACAGTGTCTATGAA 4612
QY 3874 TACACATGAAAGCAAGTCTTTTGGAAACAATAACGAACGAACTGTATGTTGTCAGCGTTCA 3933
```

```
Db 4613 TATAATGGATGGCCAGTGTCTTGGAAAGGGCTCTTCCCGAACAGTGTGTGTCAAAAGTCA 4672
QY 3934 GATGCGGTAAATGTACAGAGGAGTGTCTCCCTCAGGCCCGCTCTGTGCTGCCATTCCGGCAG 3993
Db 4673 GATGGTAAATGTAAACAGGGGCTGTCTTGTGTATGAGCCAGCTGTATGCCAGAGTCT 4732
QY 3994 TGCATTCCAGCTGCAGAAAACCTTTCTCTACTGTACACAGGGTGGAGTGTGTGTTGT 4053
Db 4733 TGTAAACCCACCGTGTAGTCAACCCCACTCGTACTGTAGCGAGACAAAACATGCCATTGT 4792
QY 4054 GAGAAGGCTATACAGAGATTAATGAATCAATGGTTTCTCTGGATTACTGCA----- 4105
Db 4793 GAAGAAGGTACACTGAAGTCAATGTCTTTAAACAGCCCTTGAGCAATGCACACTTATC 4852
QY 4106 -----TGAAGTACCAGGCTCAGAGATAAAAAGCTGATGTGAAAAACCTTTCTG-G 4157
Db 4853 CCGTGTGTGTATTATCCCAACCATGGAGGACAAAAGAGAGATGTGAAAACCAAGTCGGGCT 4912
QY 4158 GAAACACAGACCTGTGAATTCAAAATACATGATATTTTAAAGGA--TGGTCTCTTCAA 4215
Db 4913 GTACATCAACCAACCCCTCCAGTAAACCCAGCAGGACGCGGGAAGGACCTGGTTTCTACAG 4972
QY 4216 CCACCTTGATCCAGATGGCGAGTAAAAATTTGGGTTTATGGCGTTTCAGGTGGGCTTTT 4275
Db 4973 CCATTTGGCCAGATGGGAGACTAAAGACCTGGGTTTACGGTGTAGCAGCTGGGCAATT 5032
QY 4276 CTCATCATGATTTTCTTAATTTACTTCTTCTACCTTGTGTTTTCGAAGAGCCA---AAACCA 4332
Db 5033 GTGTACTCATCTTTTATTGTCTCCATGATTTATCTAGCTTGCAAAAGCCAAAGAACCC 5092
QY 4333 CATCAAGCACACCTCCCAACAGAAAGCTCTGACCTTAGCTTACGATGGAGACTTAGAC 4392
Db 5093 CAAAGAGCCAAACAAACCGACTGAACCTTTAACCTTAGCTATGATGGAGATGCCGAC 5152
QY 4393 ATGTAA 4398
Db 5153 ATGTAA 5158

RESULT 5
AAF30188
ID AAF30188 standard; cDNA; 6373 BP.
XX
AAF30188;
AC
XX
XX
DT 30-APR-2001 (first entry)
XX
DE Clone 21433858 encoding secreted protein SECPI.
KW SECPI; secreted protein; human; diagnosis; gene therapy; lung cancer; ss.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 235..5001
FT /*tag= a
FT sig_peptide 235..303
FT /*tag= b
FT mat_peptide 304..4998
FT /*tag= c
XX
XX WO200105971-A2.
XX
XX 25-JAN-2001.
XX
XX 20-JUL-2000; 2000WO-US019890.
XX
XX 20-JUL-1999; 99US-0144722P.
XX 29-NOV-1999; 99US-0167785P.
XX 19-JUL-2000; 2000US-00619252.
XX
XX (CURA-) CURAGEN CORP. PA
```





Db 2116 TACCACTGGCAAACTGGTCTCCCTGGGCGCAGTGCATTTGAGGACACCTCAGTATCGTCTTC 2175  
Qy 1624 AATGCAACCAATTTGGCTGGAATGGAGAGACCAACGCTGGTGTAGGCAATTCAGACTCGGAGA 1683  
Db 2176 AACCAACTACCACTTGGATGGGAGGCGCTCTCTCTGTGGCATGCGAGCAAGAAA 2235  
Qy 1684 GTCCTTCTGTGTCAGAGATCAGTGGGACAAGTAATGACCAAAAGATGTCAGATTCTACT 1743  
Db 2236 GTCATCTGTGTGCGAGTCAATGTGGGCAAGTGGGACCCAAAATAATGTCCTGAAGGCGCT 2295  
Qy 1744 CGACTGAACTGCGCCCTGTTTCTCCATGCAAAAGACGTATGTTGACTGCT 1803  
Db 2296 CGACTGAACTGTAAGCCCTGTCTCTCTCTTGTAAAGAGACTGTATGTGACCCCA 2355  
Qy 1804 TTCAGTGAGTGACACCCCTGCCCAGGATGTGCCAAGCAGGAAATGCCACAGT---AAAA 1860  
Db 2356 TATAGTGACTGACATCATGCCCTCTCTCTGTGTAAAGAGGGACTCCAGTATCAGGAAG 2415  
Qy 1861 CAGTCTGATACAGAAATCATCATCAAGAGCAGCCAAATGAGGCCAGGAATGCCGAGAT 1920  
Db 2416 CAGTCTAGGCATCGGGTCTCATTCATTCAGCTGCCAGCCAAACGGGGGCCGAGACTGCACAGAT 2475  
Qy 1921 ACCTTATATGAGGAGAGAGAGTGTGAAGATGTTCTCTGTCTGTATATCGTGGAG 1980  
Db 2476 CCCCTCTATGAAGAGAGGCTGTGAGGCACCTCAAGCGTGCCAAAGCTACAGTGGAG 2535  
Qy 1981 CCACAGAAATGAGCCCTTGATCTTATAGTGCAGAGTCTGTCTGGCAGGAAATAACGGGC 2040  
Db 2536 ACTCACAAATGGCCAGATGCCAATTAATAGTCCCTTTGGAGCGGTGCAACAGACAGCCCTGGA 2595  
Qy 2041 AGCAGTGAAGCCTGTGGAAGGGGTTTAAACAGAGCTGTCTCATGCTCTGTATGAC 2100  
Db 2596 GCACAGGAGGCTGTGGCCCTGGCGCAGCAGCAGAGCCATTACTTGTCCAGCAGAGAT 2655  
Qy 2101 AACCGGTACAGAAATGATGGAAATGCTCAAGCAGACAAACGCGATGCCCTCTCTGTG 2160  
Db 2656 GGAGCAGAGGCTGGAATCCATGAGTGGCTACAGTATGCAGGCCCTGTGCAGGCCCTTACC 2715  
Qy 2161 CAAGAAATGCACAGTCCCATGTCGAGAGAGCTGCACCTTCACTGCTGGTCCAGTTTACG 2220  
Db 2716 CAGGCTGCCAGATCCCTGCGCAGATGACTGTCAATTGACCGAGCTGGTCCAGTTTCT 2775  
Qy 2221 CCCTGCTCCAGAACTGTGAAGCCACAAAAGTAGTGGCGGCGACAGCTCAACAGGAAAAGC 2280  
Db 2776 TCATGCAATGAGACTGTGTGCTAGTTAGGACAGAAAAGCGACTCTTGTGGAAAAGT 2835  
Qy 2281 AGAAGAGAGGAATGCCAGATCTTGACCTTTACCTTTAGTGGAGACAGACTATGT 2340  
Db 2836 AAAAGAGGAAAATGTAAAAATTCACATTTGTATCCCTGTATGAGACTCAGTATTGT 2895  
Qy 2341 CTTTGTGATGAATTTATATCCCAACCTTATGGAACTGGTCAGATTGCATTTCTCCAGAA 2400  
Db 2896 CTTTGTGACAAATATATGACACACCTCTGGGAACTGGTCAGACTGTATTTTACCAGAG 2955  
Qy 2401 GGCAGAGGAGCCTCACCGAGAGCTGGGGTACAAGCAGACAGCAAGAAATGTGGAGAA 2460  
Db 2956 GGAAGAGTGAAGTGTCTGGGAATGAAGTACAAGGAGACATCAAGGAATGCGGACAA 3015  
Qy 2461 GGCCTGCGCTTTTCAGAGCAGTAGCTGTTCTGATAAAAATGAAGACCTGTGACCCCTCC 2520  
Db 3016 GGATATCGTTACCAAGCATGCGATGCTACGATCAAAATGGCAGCTTGTGAAAACATCT 3075  
Qy 2521 TTTGCGACAGCTGTGTTACATTAAGAAAATGTGTGATTCCTGCCCATTTGATTGC 2580  
Db 3076 AGATGTAACAGCCATGGTTACATTTAGAGAGCCCTGCATCATCCCTGCCCTCAGACTGC 3135  
Qy 2581 AAGTTAAGCAGTTGGTCTAGTTGGGCTCTTGAGTTTCATCTTGTGAAATGGAGTGAGA 2640  
Db 3136 AAGCTCAGTGAGTGGTCCAACTGGTCCGCTGCGAGCAAGTCCCTGTGGGAGTGGTGAAG 3195  
Qy 2641 ATTCGATCCAAATGGCTAAAAAGAAAAACCTTTACAAATGGAGGACGACCATGTCCCAAACTG 2700

Db 3196 GTTCGTTCTAAATGGCTCGTGAAAAAACCATATAATGAGGAGAGGCCCTTGCCCAAACTG 3255  
Qy 2701 GATC-----TCAGAAATCAGSTACAGCAGTCCCATGTTTACGTGAGTGAATCAG 2754  
Db 3256 GACCATGTCAACAGGCAAGGTATAGGTTGTCCCATGCCACAGTGAATGCAACCCAG 3315  
Qy 2755 TATTCCTGGTGTAGAACACTGGTCTTCATGCAAAATCAACAATGAGCTGAGTCCCTG 2814  
Db 3316 TACCTATGGGTCAACAGAGCCCTGGAGCATCTGCAAGGTGACCTTGTGAATATGCGGAG 3375  
Qy 2815 CGCTGTGAGGAGGAGAACCAATCTAGGAAATCAGATGTGT---GAATACTGCGGATGGT 2871  
Db 3376 AACTGTGAGAGGGCGTGCAAAACCCGAAAAGTGAGATGCATGCGAAATACAGCAGATGCG 3435  
Qy 2872 GAAGTGGAGCAGTGGATAGCAACCTGTGCAACAGGATGAAATTCGCCCAAGAAACCCAG 2931  
Db 3436 CCTCTGAACATGTAGAGGATTACCTCTGTGACCCAGAGAGATGCCCTGGCTCTAGA 3495  
Qy 2932 TCCTGTTCTTATGTGTGCCAATGAGTGTGTCTGATCTGAGTGGGACCTTGGAGCAAA 2991  
Db 3496 GTGTGCAAAATACCATGCCCTGAGGACTGTGTGATATCTGAATGGGTCCATGGAACCAA 3555  
Qy 2992 TGCCCAACAGTCAATGCGATCCCAACAAAT---GCAGAGAAAGAACTGCGCACTGTAGA 3048  
Db 3556 TGTGTTTGGCTTGCATCAAGCAGTTTCGCGCAAGGTGAGTCCATCCATCAGACAA 3615  
Qy 3049 CCATCACTGAACCTAAGGACTGTGTGAGACTCAACAGTGCAGCCCTTGCTCTCTGAT 3108  
Db 3616 CCAGCTGATGAAGGAAGATCTTGCCCTAATGCTGTGAGAAAGAACCCCTGTAACCTGAAC 3675  
Qy 3109 GAAAATTCCTCCAGTTCAGGTACAACTTAACAGAGTGGACACATGCCAGCTCAGTGA 3168  
Db 3676 AAAAATCTCTACCTATGATTAATGTAACAGCTGGAGTACATGTACGTGAGTNG 3735  
Qy 3169 AACGACCCCTGTGTGCAAGCGCTCAGGACCCGCTGTGCTAAGCTGTGTGTGAGTGTG 3228  
Db 3736 AAGGAGTGTGTGGAATGGAATAAAACAAGGATGTGGATTTGTGTGGAAGTGTG 3795  
Qy 3229 AAGCAGTCAGATGAGCAATGTGNGCAGCATAAATTTGGAGAGAGCCCAAGAGATGAGC 3288  
Db 3796 AAGTCAAGTGTGCAATTAATTTGGAAGCGCTTGGCTTGGAGAGAACTGGCAGATGAAC 3855  
Qy 3289 APTCCCTGCTTGTGTGAAATGGTGGTCAACTGTGCTCAGCTCTCAGGGTGGACCGCTTGA 3348  
Db 3856 AGTCTCTGATGTGGAATGCCGTGAACTGTGAGCTTTCTGATTGCTCTCTGTTGTTCA 3915  
Qy 3349 GAGTGTTCACAGACCTGTGTGGCATGGAGTGAATGAGCCGAGTCTGATTTATCATATG 3408  
Db 3916 GAATGTTCTCAACATGTGGCTCACAGGAAAAATGATCCGAAAGACGAACTGACCCAG 3975  
Qy 3409 CCAACCCAGAGAGAGGAGCGCCATGCCCAACAGAGCTTTACCCAGGAGAAACCTGCCCCA 3468  
Db 3976 CCGTTTCAAGGTGAGGAAGACCATGCCCTTCCCTGATGGACAGTCCAAACCTGCCCCA 4035  
Qy 3469 GTGACCCCTGCTACAGCTGGGTCTTGGCAACTGTCTGCAATGTAATTTGGAGGTTGGA 3528  
Db 4036 GTGAAGCCCTTGTATTCGGTGGCAATATGSCCAGTGTCTCCATGCAAGTGCAGAGGCC 4095  
Qy 3529 GACTGTGGGAGGAGTTTCAGATCCGAGCCTTTCTGATGTTCCATGTGTTCCACAGTGGTTCAATA 3588  
Db 4096 CAGTGTGAGAGAGGAGCCAGAACAGGAACATTTCTGTGTAGTGAATGAGTGGTCACT 4155  
Qy 3589 TCTCATGAGCTGGAGCTGTGAGGATGCACTGTGTGGAGAAATGCCCTT----- 3638  
Db 4156 GATGATTTTCAAGAAAGTGTGGAATGAGGAATTTCTGTCTGACATTTGAATCATTTATAGAT 4215  
Qy 3639 -----TCAGGACAGCATCCTGAACAGCTGTGTTCTGTGCTTCCAGGAGACTGCCAT 3693  
Db 4216 GGTAAATAAATAATGTTCTGGAGGAATTCCTGACGCCAGCCCTTGCCCAAGTGACTTTAT 4275  
Qy 3694 TTAACAGAAATGCTCAGAGTGGAGCAGATGTGAATTAACCTGCATTTGATGGAGAGCTTT 3753  
Db 4276 TTGAAGGACTGCTCTCTCTGGAGCCTGTGTGAGCTGACCTGTGTGAATGGTGGAGATCTA 4335



Db 685 CATGTGTGGGCCCCCGCGAGTTCCGAGGCTCTGCTGTCTCAAACTGTACGAGATTCCAG 744  
Qy 250 GCCTGTGATGCTCCCATTTCTCTGTCCTCTTGGGAGAGAGAAATATACATTTAGCCCTTAAG 309  
Db 745 GTGTG-----CCAATCCAGTCCATGCGAGCGCGAGAGTCAAGTACAGCCTGCAAT 795  
Qy 310 GTTGGACATGGAGTAATGACAGACTGCTCTCATCTTAAAGAAATTAATCCAAGCGGAAGA 369  
Db 796 GTGGGGCCCTGGAGCACTGCTCAATGCCCACTCCCGCAAGTAAGACAAAGGAGA 855  
Qy 370 ACTGTTCTGGATTTTAA-----CCTGATTCAAAATGAG 402  
Db 856 CGCGGGAAGATAAAGAACCGGGAAGACCGCGAGCAAGAGTAAGGATCAGAGCC 915  
Qy 403 CGAGTCACTTTAAACATCAAGTTACA-----AAGCATACTATCTCGAAGTCT 453  
Db 916 CGCGAGCTTAAAGAAAGAGAAACAGAAACAGGAGCAAGACAGAAAGAAACAATAT 975  
Qy 454 TGGGCAATAGAGATAGGTTATCAAAACCGGCGAGGTTCTGTTACAGAAAGTATGACAA 513  
Db 976 TGGACATCCAGATTGGATATCAGACCAGAGAGGTTATGTGCATTAACAAGCGGGAAA 1035  
Qy 514 AATGCTATGTTAAGCCTTTGCTTTCAAGATTCTTCCCATTTGACTGTTCAGTCTCTGATC 573  
Db 1036 GCTGCTGATTTAAGCTTTTGGCAAGAAAGAGAGCTTCCAATGACCTTCCAGTCTCTGTG 1095  
Qy 574 ATGCCCAAAGACTGTGAACCTCCAGTGTCTCTCGAGAGCCCTGCTCCAAAGACATGC 633  
Db 1096 ATCAACAAAGAGTCCAGGTTCCGAGTGTTCAGAGTGGAGCCCTGCTCAAAAACATGC 1155  
Qy 634 CGTTCAGGAGTCTCTTGGCAGGATTTAGAGCAGGAGCGCGAAGCACTGAGCACTGGCT 693  
Db 1156 CATGACATGTTGTCCTCGAGGCACTCGTGTAAAGACAGAACCATCAGCAGTTTCCC 1215  
Qy 694 ATTGGAGTGGAAAGAGTGTCTCGAATCTTTCAGAAAGAGGCGCTGCATTGTGAGGA 753  
Db 1216 ATTGGCAGTGAAGAGAGTGTCCAGAAATTTGAAGAAAAGAACCTGTTTGTCTCAAGGA 1275  
Qy 754 GAACTTCTGCAGCAATCTCCAGGATTTCTCGGAGAACTTCTGAATGGAAGAAATGCCAA 813  
Db 1276 GATGGAGTTGTCCTGTCGACGATGCTGTGGAGAACTACAGAGTGGAGTGCCTG 1335  
Qy 814 GTCTCTCTCTCTCGAGCAGCAGGATCCCACTGGGATGTGAGGAGCCGTGTGTGGC 873  
Db 1336 GTGACCCCTTGTCTAGTCAAGCAGGACAAAGAGCGCGGCAACAGACGCGCCCTCTGTGA 1395  
Qy 874 GGTGGATCCAGACCCGGAGGTGTACTGTGCCAG-----AGCTACCAAGCA 921  
Db 1396 GGGGGCATCCAGACCCGAGAGGTGTACTGTGCGAGGCCAACGAAACCTCTCTCACAA 1455  
Qy 922 GCTCCGCACTGAGGGCCAAAGAAAGTCTTAGACCTGTGGAAGGCAATTTATGTGGGA 981  
Db 1456 TTAAAGTACCACAAAGAAAGAGCCTCAAAGCCAATGGACTTAAATTTATGCACCTGGA 1515  
Qy 982 CCCGCCCGTTGCCCTCTAGCTCTGCAATATCCCTTGTCTACGAGCTGCATAGTATCT 1041  
Db 1516 CCTATCCCTAATACTACAGACTGTGCCATCTCTTGTCCAACTGAATGTGAAGTTCA 1575  
Qy 1042 TCCTGGTCAGCCTGGGCGCTGTGCATCCATCAAAACTGTCAAGCCTCAGGGGAAAAA 1101  
Db 1576 CTTTGGTCAGCTGGGACCTTGTACTTATGAAACTGTAACTGATCAGCAAGGGAAGAAA 1635  
Qy 1102 GGATTTAGNACGAGCGAG-----CCATGTCCTCATGNACTACAGGCGCTGCA 1152  
Db 1636 GGCCTTCAAACTGAGGAAGCGGCGCATTTACCAATAGAGCCCACTGGAGGCTCTGGGTAA 1695  
Qy 1153 GGGCAATTTGCCCTCATTTTGGTGGAGTCTGTTCTTGTGAGGATCCAAATGTCTACCGATGG 1212  
Db 1696 GAAAACAGCCCTCACTTACTTGAAGAGCCATCTCCCTGTGAGAGCGCTGCTGTATGACTGG 1755  
Qy 1213 ---CTGGCATCAGAGGATCTGTTTCCCTGTATCATGGAA---AATGTGCCCTGGGACAT 1266  
Db 1756 AAACCGGTGAGCTGGGAGACTCGGAGCCAGATAACGGAAGAGAGTGTGTCCAGGCAGC 1815

Qy 1267 CGTATTTCTGAAGCGCGTGTGCCAGATGACCGCGAGAAAGATGTATCAGGAGTCTTTGC 1326  
Db 1816 CAAGTTCAAGAGGTTGTGTGCATCAACAGTATGAGAGAAAGTTGACAGACACTGTGC 1875  
Qy 1327 CCAGTTCCCTCTCTCTGAGAGGAAGTCTTGTGAAATTCCTTCGCGAATGGACTGTG 1386  
Db 1876 AGAGATGCCATCTTCCCATCCCTGTGCGCTGTATGCCCATGCCGAAAGACTGTGTG 1935  
Qy 1387 GTGACGAGTGGAGCGGAGTGTGTATCTCTGTTCCTCCAGTCTCTTCAAAATAAAATCTCAGAT 1446  
Db 1936 CTCAGCACATGCTCTACGTGTCTCTCTGTCTCAACACCTGTCTCAGGGAACACAGAA 1995  
Qy 1447 GGGAAACAGACAGGTCAAGAACTATCTCTGSC-----ACTGGCTGGGAGGTGGAAAGCCA 1503  
Db 1996 GGGAAACAGATACGAGCACGATCCATTTCTGSCCTATCGGGTGAAGAGGTGGAATTCGC 2055  
Qy 1504 TGTCCCTCTAGTCAAGGCTCTCCAAAGAGCATCGTTTGTGTATGACCATTCCTGTATGCAG 1563  
Db 2056 TGTCAAAATGCAAGTCTTTCAGAGAGTAGAAGCTGTATGAGCATCTTGTGCACAGTG 2115  
Qy 1564 CTTACTGAGAGACATCGCTTGGGGCCCTTGTCTGAGAGACATTTGGTAACTGCGCCTT 1623  
Db 2116 TACCACCTGCAAACTGTGTCTCTGGGGCCAGTGCATTTGAGGACACTCAGTATCTGCTTC 2175  
Qy 1624 AATGCAACCATTTGGCTGGAATGGAGAGCCACGCTGTGTAGGCATTCAGACTCGGAGA 1683  
Db 2176 AACACAACTACGACTTGGAAATGGGAGGCTCTCTGTCTGTGCGCANTGCAGCAAGAAA 2235  
Qy 1684 GTCTTCTGTCTCAAGAGTCACTGCGGAACAATATGACCAAAAGATGTCCAGATTTCTACT 1743  
Db 2236 GTCACTCTGTGCGAGTCAATGTGGGCCAAGTGGGACCCAAAAAATGCTCTGAAAGCCTT 2295  
Qy 1744 CGACTGTAATCTGTGCGCCCTGTTTCTCCATGCAAAAGAGCTGTATTTGTCAGTCT 1803  
Db 2296 GCACTGTAATCTGTAAAGCCTTGTCTCTTGTAAAGAGGACTGTATTTGACCCCA 2355  
Qy 1804 TTCAGTGTGAGCACCCCTGCCCAGAGATGTCCAAGCAGGAAATGCCACAGT---AAAA 1860  
Db 2356 TATAGTACTGGACATCATGCCCCCTCTTCGTGTAAAGAGGGGACTCCAGTATCAGGAAG 2415  
Qy 1861 CAGTCTCATACAGNATCATCTCAAGAGACAGCCAAATGGAGGCCAGGATGCCCAGAT 1920  
Db 2416 CAGTCTAGCGCATCGGGTCAATTCAGCTGCCACCAACGGGGGCCGAGACTGCACAGAT 2475  
Qy 1921 ACCTTATATGAGGAGAGAGAGTGTAAAGATGTTTCTTGTCTCTGTATATCGTGTGAAG 1980  
Db 2476 CCCCTCTATGAAGAGAGGCTGTGAGGCACCTCAAGCGTGCCAAAGCTACAGTGGNAG 2535  
Qy 1981 CCAAGAAATGAGGCCCTTGTGATCTTATGTGCCAGAGTCTGTCTGGCAGGGAATTAACGGGC 2040  
Db 2536 ACTCAAAATGGCGCAGATGCCAAATTTAGTCCCTTTGGAGCGGTGCAACAGACAGCCCTGGA 2595  
Qy 2041 AGCAGTGAAGCCTGTGGAAGGGGTTACAAACAAGAGTGTCTCATGCTCTCTGATGAC 2100  
Db 2596 GCAAGGAGAGCTGTGGCCTGGGCGACAGCAGAGGCCATTACTTGTTCGCAACAGAT 2655  
Qy 2101 AACCGGTGAGCAGAAATGATGGAATGCTCAAGCAGACAAACGGCATGCTCTCTCTGTG 2160  
Db 2656 GGAGACAGAGCTGGAATCCATGAGTGTGCTACAGTATGAGGCGCCCTGTGCCAGCCCTTACC 2715  
Qy 2161 CAAAGATCCACAGTCCCATGTGCGAGAGACTGCCACCTTCACTGCTGTCTCAAGTTCACG 2220  
Db 2716 CAGGCTGCCAGATCCCTCTCCAGAGTACTGTCAATTTGACAGCTGGTCCAGTTTCT 2775  
Qy 2221 CCCTGCTCCAGAACTGTGAAGCCACAAAAAGTAGTGGGGGACAGCTCACAGGGGAAAGC 2280  
Db 2776 TCATGCAATGAGACTGTGTGCGATTTAGGACCAAGAAAGCGCACTCTTGTGGAAGAGT 2835  
Qy 2281 AGAAGAGAGGAGAAATGCCAGGATTTCTGACCTTTACCTCTAGTGGAGACAGACTATGT 2340  
Db 2836 AAAAGAGAGGAAAAATGTAAAAATTTCCCAATTTGATCCCTGATTTGAGACTCAGTATTGT 2895

2341 CCTTGTGATGAATTTATATCCCAACCTTATGGAACCTGTCAGATTGCATTTCTCCAGAA 2400  
2896 CCTTGTGACAAATATATATGCACAACCTGTGGGAACTGTCAGACTGTATTTTACCAGAG 2955  
2401 GGCAGAGGAGGACCTCACCGAGACTGCGGTTACAAGCAGACAGCAAGAAATGTGGAGAA 2460  
2956 GGAAGAGTGGAGTGTGCTGGGAATGAAAGTACAGAGGACATCAAGGAATGCGGACAA 3015  
2461 GGCCTCGCTTTTCGAGCAGTAGCTTTCTGATATAAAATGGAAGACCTGTGACCCCTCC 2520  
3016 GGATATCGTTACCAAGCAATGGCATGCTACGATCAAAATGGCAGGCTGTGGAACATCT 3075  
2521 TTCTGACAGCAGCTCTGGTTACATTAAGAAAAATGTGTCAATTCCTGCCCATTTGATTCG 2580  
3076 AGATGTAACAGAGCCATGGTTACATTTGAGAGGCTGTCATCATCCCTGCCCTCAGACTGC 3135  
2581 AAGTTAAGCGATTGGTCTAGTTGGGGCTTTGCGAGTTCACTTTGTGGAAATGAGAGTGAGA 2640  
3136 AAGCTCAGTGAGTGTCCAACTGGTCCGCTGCAGCAAGTCTGTGGAGTGGTGTGAAG 3195  
2641 ATTTCGATCCAAATGGGTAAAGAAAAACCTTACAATGGAGGACGACCAATGTCCTCCAACTG 2700  
3196 GTTCGTTCTAAATGGCTGGTGAAGAAACCATATAATGGAGGAAGGCTTGTCCCCAACTG 3255  
2701 GATC-----TCAGAAATCAGGTACATGAGGAGTCCCATGTTACAGTGTGCAATCAG 2754  
3256 GACCATGTCAACAGGACAGGTGTATGAGGTTGCTCCCATGCCACAGTACTGCAACCCAG 3315  
2755 TATTCCTGGGTTGTAGAACACTGGTCTTCATGCAAAATCAACAATGAGCTGAGTGCCTCG 2814  
3316 TACCTATGGGTACAGAGCCCTGGAGCATCTCAAGGTGACCTTGTGAATATGCGGAG 3375  
2815 CGCTGTGGAGGAGGACAACTTAGGAAATCAGATGTGT---GAATACTCGGATGGT 2871  
3376 AACTGTGGAGAGGCGGTGCAACCCGAAAGTGTGATGATGATGATGATGATGATGATGATG 3435  
2872 GAAGGTGAGCAGTGTAGTACCACTGTGCAACAGGATGAAATTCCTCCAGAAACCCAG 2931  
3436 CCTTCTGACATGTAGAGGTTACCTCTGTGACCCAGAGAGATGCCCTGGCTCTAGA 3495  
2932 TCTGTCTTCTTATGTGTCCCAATGATGTGTATGTGTGATGGGAGCTTTGGAGCAA 2991  
3496 GTGTGCAAAATTACCATGCTGAGGACTGTGTATATCTGAAATGGGCTCCATGGACCAA 3555  
2992 TGCCACAGTGTGATGCTCCACACAAAT---GCAGAGAAATCGCCACCTGTCTAGA 3048  
3556 TGTGTTTGCCTTGCATCAAGAGCTTTCCGCAAGGTGAGTGTGATCCCATCAGACAA 3615  
3049 CCATCACTGAACCTCAAGGACTTTGTGTGAAGACTCACAGGTGAGCCTTGCCTCTGAT 3108  
3616 CCAGCTGATGAAGGAAGATCTTGGCCCTAATGCTGTTGAGAAAGACCTGTAACTGAAC 3675  
3109 GAAATTTGCTTCCAGTTCAGTACATCTAACAGAGTGGAGCAGATGCCAGCTGATGAA 3168  
3676 AAAACTGTCTACCATATGATTTATATGTAAGTAAAGAGTGTGATGATGATGATGATGATG 3735  
3169 AAGCAGCCTGTGGTCAAGGCGTCAGGACCCGCTGCTAAGCTGTGTGTGAGTGTGATGCG 3228  
3736 AAGGCAAGTTTGGAAATGGAATATAAACAAGATGTTGGATTGTTGCAAGTGTATGCG 3795  
3229 AAGCAGTGTGAGCAACATGTGAGCAGCAATATTGGAGAACCCCGAGAAATGAGC 3288  
3796 AAGTCAGTTGACCTGAAATATTGTGAAGCGCTTGGCTTGGAGAAACTGGCAGATGAAC 3855  
3289 ATTTCCTGCTTGGTGAATGCGTGTGCACTGTGAGTCTCAGGCTGAGCGCTTGGACA 3348  
3856 ACGTCTGATGTTGGAAATGCGCTGTGAATCTGTGAGCTTCTGTATGTTGCTTCTGGTCA 3915  
3349 GAGTGTTCACAGCCTGTGGCCATGAGGTGCAATGAGCCGAGCTCGATTATCATATTATG 3408  
3916 GAATGTTCTCAAAACATGTGGCTTCAACAGGAAATGATCCGAAAGACGACAGTGTGCCAG 3975  
3409 CCAACCCAGGAGAGGAGCGGCCATGCCCCCAGAGCTTTACCCAGGAGAAACCTTGCCCA 3458

Db 3976 CCCTTTCAAGGTGATGGAAGACCATGCCCTTCCTGATGAGCAGATCCCAACCTGCCCA 4035  
QY 3469 GTGACCCCTCTGTACAGCTGGGTCTTGGCAACTGGTCTGCATGTAAATTTGAGAGGTGGA 3528  
Db 4036 GTGAAGCTTTGTTATCGGTGGCAATATGGCCAGTGGTCTCCATGCCAAGTGCAGGAGGCC 4095  
QY 3529 GACTGTGGGGAAGAGTTCAGATCCGAGCCTTTCCCTGCAATGCTCCACAGTGGTTCATA 3588  
Db 4096 CAGTGTGGAGAGGGGACCAGAAACAAGAACATTTCTTGTGTAGTAAAGTGTGGTCACT 4155  
QY 3589 TCTCATGCACTGGAACGTGTGAGGATGCACTGTGAGGAGAAATGCCCTT----- 3638  
Db 4156 GATGATTTCAAGCAAGTGGTGGATGAGAAATCTGTGCTGACATTTGAATCAATATAGAT 4215  
QY 3639 ----TCAGAGACAGCATCTTGAAGCAGCTGTGTTCTGTGCTTGGCCAGGAGACTGCCAT 3693  
Db 4216 GGTATATAAAATATGTTCTTGGAGGAATCTCTGAGCAGCCTTGGCCAGGTGACTGTAT 4275  
QY 3694 TTAACAGAAATGTCAGAGTGGAGCAGATGTGAATTAACCTGCAATTCATGGAAGAAGCTTT 3753  
Db 4276 TTGAAGGACTGGTCTTCTCGAGCCTGTGTGCTGACCTGTGTGATGCTGAGGATCTA 4335  
QY 3754 GAGACTGTGGCCCGCAGTCTAGATCAAGCACTTTTATATTCAGTCTTTTGAAGACCA 3813  
Db 4336 GGCTTTGGTGGAAATACAGGTCTGATCCAGACCGGTGATTATACAAGAACTAGAGAAATCAG 4395  
QY 3814 GACAGCTGCCCCCAACAGAGTCTTAGAAACACGCGCTTGTACAGAGGCAATGTATTCAC 3873  
Db 4396 CATCTGTGCCCAGAGCAGATGTTAGAAACAATCATGTTATGATGACAGTCTATGAA 4455  
QY 3874 TACATGTGAAGAGCAAGTCTTTTGGAAACAATAACAGAACTATGATGGTCCAGCGTTCA 3933  
Db 4456 TATAATGAGTGGCCAGTGTCTTGGAAAGGCTCTTCCGAAACAGTGTGGTCTCAAGGTCA 4515  
QY 3934 GATGGCTTAATGTCAAGAGAGCTGCTCCCTCAGCCCGTCTCTGCTGCAATTCGSCAG 3993  
Db 4516 GATGTTATAATGTACAGAGGAGTCTGCTGCTGATGAGCAGCCTGATGCCAGAGTCT 4575  
QY 3994 TGCAATCCAGCCTGCAGAAACCTTTCTCTACTGTACAGAGGTGGAGTCTGTGGTGT 4053  
Db 4576 TGTAAACCCACCGTGTAGTCAACCCCACTGCTACTGTAGCCAGACAAAAACATGCCATTGT 4635  
QY 4054 GAGAGGCTTATACAGAGATAATGAATCAAAATGGTTCCTGGAATTAAGTCA----- 4105  
Db 4636 GAGAGGGTACACTGAAGTCAATGTTCTTAACAGCACCCCTTGAGCAATGACACTTATC 4695  
QY 4106 -----TGAAAGTACCCAGGCTCAGAGGATAAAAAAGCTGATGAAAAACCTTTCTG-G 4157  
Db 4696 CCGTGGTGTGTTATCCCACTAGGAGGACAAAGAGGAGATGTGAAACCACTCGGCT 4755  
QY 4158 GAAAAACAGACCTGTGAAATTCAAAAATAATGATATTTTTTAAAGGA--TGGTCTCTTCAA 4215  
Db 4756 GTACATCCAAACCAACCCCTCCAGTAAACCCAGAGGAGGAGGACCTGCTTTCTACAG 4815  
QY 4216 CCATCTGATCCAGATGGCCGAGTAAATAATTTGGGTTTATGGCGTTTCAGGTGGCGCTTTT 4275  
Db 4816 CCATTTGGCCAGATGGGAGACTTAAGACCTGGGTTTACGGTGTAGAGCTGGGGCATTT 4875  
QY 4276 CTATCATATGATTTTCTTAAATTTTATCTTCTTCTGTTTTCAGAGAGGCA---AAACCA 4332  
Db 4876 GTGTTACTCATCTTTATTTGTTCTCCATGATTTATCTAGCTTGTCAAAAAGGCAAAAGAACCC 4935  
QY 4333 CATCAAGACACCTCCCAACAGAGGCTCTGACCTTACCTACCTGAGACTTAGAC 4392  
Db 4936 CAAAGAGGCAAAACCAACCGACTGAAACCTTTAACTTACCTTATGATGAGAGATGCCGAC 4995  
QY 4393 ATGTAA 4398  
Db 4996 ATGTAA 5001

RESULT 7

ACC72171  
ID ACC72171 standard; DNA; 6373 BP.  
XX  
AC ACC72171;  
XX  
DT 07-JUL-2003 (first entry)  
XX  
DE Human NOV47j coding sequence.  
XX  
KW Human; NOV; antidiabetic; anorectic; antibacterial; virucide;  
KW immunomodulator; cytosolic; nootropic; neuroprotective;  
KW antiparkinsonian; antilipemic; gene therapy; metabolic disorder;  
KW diabetes; obesity; infection; cachexia; cancer; dyslipidaemia;  
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
KW immune disorder; haematopoietic disorder; gene; ds.  
XX  
OS Homo sapiens.  
XX  
PN W02003029423-A2.  
XX  
PD 10-APR-2003.  
XX  
PF 02-OCT-2002; 2002WO-US031358.  
XX  
PR 02-OCT-2001; 2001US-0326483P.  
PR 05-OCT-2001; 2001US-0327342P.  
PR 09-OCT-2001; 2001US-0327917P.  
PR 09-OCT-2001; 2001US-0328029P.  
PR 09-OCT-2001; 2001US-0328044P.  
PR 09-OCT-2001; 2001US-0328056P.  
PR 12-OCT-2001; 2001US-0328849P.  
PR 15-OCT-2001; 2001US-0329414P.  
PR 17-OCT-2001; 2001US-0330142P.  
PR 22-OCT-2001; 2001US-0341058P.  
PR 24-OCT-2001; 2001US-0339266P.  
PR 24-OCT-2001; 2001US-0343639P.  
PR 29-OCT-2001; 2001US-0349575P.  
PR 01-NOV-2001; 2001US-0346357P.  
PR 12-APR-2002; 2002US-0371972P.  
PR 12-APR-2002; 2002US-0371980P.  
PR 17-APR-2002; 2002US-0373261P.  
PR 19-APR-2002; 2002US-0373805P.  
PR 23-APR-2002; 2002US-0374738P.  
PR 16-MAY-2002; 2002US-0381101P.  
PR 17-MAY-2002; 2002US-0381635P.  
PR 29-MAY-2002; 2002US-0383830P.  
PR 01-OCT-2002; 2002US-00262839.  
XX  
FA (CURA-) CURAGEN CORP.  
XX  
PI Alsbrook JP, Anderson DW, Boldog FL, Burgess CE, Catterton E;  
PI Edinger SR, Ellerman K, Gerlach VL, Gorman L, Guo X, Ji W;  
PI Kekuda R, Leach MD, Li L, Miller CE, Patturajan M, Rieser DK;  
PI Rothenberg ME, Shinkens RA, Smithson G, Spytek KA, Taupier RJ;  
PI Vernet CAM, Voss EZ, Zerhusen BD, Zhong M;  
XX  
DR WPI; 2003-381625/36.  
DR P-2SDB; ABR59459.  
XX  
XX NOVX polypeptides and nucleic acids useful for diagnosing, preventing or  
XX treating NOVX-associated disorders, e.g. diabetes, obesity, cancer or  
XX dyslipidemia, and in chromosome mapping, tissue typing or  
XX pharmacogenomics.  
XX  
PS Claim 20; Page 263-264; 487pp; English.  
XX  
CC The present invention relates to novel human NOV proteins and their  
CC coding sequences (ACC72075-ACC72181 and ABR58363-ABR58469). The NOV  
CC proteins are useful in manufacturing a medicament for treating a syndrome  
CC associated with a human disease. The NOV proteins and coding sequences  
CC may be used to diagnose, treat or prevent metabolic disorders such as  
CC diabetes or obesity, infections, cachexia, cancer, neurodegenerative  
CC disorders such as Alzheimer's disease or Parkinson's disease, immune

CC disorders, haematopoietic disorders and various dyslipidaemias  
XX  
SQ Sequence 6373 BP; 1836 A; 1385 C; 1603 G; 1548 T; 0 U; 1 Other;  
Query Match 28.0%; Score 1232.8; DB 7; Length 6373;  
Best Local Similarity 57.7%; Pred. No. 0;  
Matches 2598; Conservative 0; Mismatches 1782; Indels 126; Gaps 17;  
QY 10 TGCATTTCAGAGCTGAACCGAACTGTGGTTCGAATGAAATATCGAACACTTTGCCCTT 69  
DB 505 TCATCCAGAAAGACAAAGACATTCCTCGGAGATATCATCTGTAGTATTTGAGCCC 564  
QY 70 CAGCCTCTACAGACAGCTTGCTCATTCCTTGTCCCGGATTTGTAGTATCTGAG 129  
DB 565 AAGCCTCTCTGGAGAGCTTGCTCATTCCTTGTCCAGCAAGATTGATCTGTCTGAA 624  
QY 130 TTCTTACCATGTCACACTGTAGCAAGGATGTGGAGAAATTCGAGCATAGACTGC 189  
DB 625 TTTTCTGCTGTGTCGAATGCTCAAGACCTGCGGACGGGCTCCAGCACCGGACGGT 684  
QY 190 GCGGTCTATAGCTCCCTCTCTTTTGGTGGTTTGCAATGTCCAAATCTGACTGAGTCAAGA 249  
DB 685 CATGTGTGGCGCCCGCCAGTTCGGAGGCTCTGGCTGTCCAACTGACGAGTTCAG 744  
QY 250 GCCTGTGATGTCCTCATTCCTGTCTCTTGGGAAGAGGAATATACATTTAGCTTAAG 309  
DB 745 GTGTG-----CCAAATCCAGTCCATGCGAGCCGAGGAGCTCAGTACAGCCTGCAT 795  
QY 310 GTTGACCATGGAGTAAATGCAGACTGCTCATCTTAAAGAAATTAATCCAAAGCGGAAGA 369  
DB 796 GTGGGCTCTGGAGCAGCTGCTCAATGCCCACTCCGCAAGTAAGACAAGCAAGGAGA 855  
QY 370 ACTGTTCTGGATTTTAA-----CTCTGATTCAAAATGAG 402  
DB 856 CGCGGGAAGAATAAAGAACGGGAAAGGACCGCAGCAAGAGGTAAGAGATCCAGAGGCC 915  
QY 403 CGAGTCACCTTTAAACATCAAAAGTTTACA-----AAGCACATCATCTTCGAAGTCT 453  
DB 916 CCGGAGCTTTATTAAGAAAGAGAAACAGAAACAGGCGAGAGAGGTTATGTGCAATTAAACAGCGGGGAAA 1035  
QY 454 TGGGCAATAGAGATAGTTTATCAAAACCGGAGGTTTCGTGTACAAGAAAGTATGAGCAA 513  
DB 976 TGGGACATCCAGATTGGATATCAGACCAGAGAGGTTATGTGCAATTAAACAGCGGGGAAA 1035  
QY 514 ATGCTATGTTAAGCTTTGGCTTCAGATTCCTTCCCATTTGACTGTTTCACTCTGATC 573  
DB 1036 GCTGCTGATTAAAGCTTTTGGCAGCAAGAGAGAGTTCCTCAATGACCTTCCAGTCTCTGTGTG 1095  
QY 574 ATGCCCCAAGAGCTGTGAACCTCCAGTGGTCTCTCGAGCCCTCTCCCAAGACATGC 633  
DB 1096 ATCACCAAGAGTGCAGGTTTCCGAGTGGTCAGATGGAGCCCTGCTCAAAAACATGC 1155  
QY 634 CTTTCAGGAGTCTCTTCCAGGATTTTAGGAGCAGGACCGGAAAGTGAAGCAATGGCT 693  
DB 1156 CATGACATGGTGTCCCTCGAGGCACTCGTGTAGGACACAGCAACCATCAGGCACTTTCCC 1215  
QY 694 ATTGAGGTGGAAAGAGTGTCTGAACTTCTTGAGAAAGAGGCTGCTATTTGGAAGA 753  
DB 1216 ATTGAGGTGAAGAGTGTCCAGATTTTGAAGAAAGAACCTGTTTGTCTCAGGA 1275  
QY 754 GAATTTCTGAGCAATGTCCAGGATTTCTTGGAGAACTTCTGAATGGAAGAAATGCCAA 813  
DB 1276 GATGAGTGTCTCCCTGTCCACGCTATGGCTGGAGAACTACAGAGTGGACTGAGTGGCGT 1335  
QY 814 GTCTCTCTCTCTCGAGCAGCAGGATCCCACTGGCATGTGAGGGACCGCTGTGTGGC 873  
DB 1336 GTGAGCCCTTTGCTCAGTTCAGCAGGAAGAAGGCGCGGCAACAGAGCCCTCTGTGA 1395  
QY 874 GGTGGATCCAGACCCGGAGGTGACTGTGCCAG-----AGCGTACCAGCA 921  
DB 1396 GGGGGCATCCAGACCCGAGAGGTGACTGCGTGCAGGGCCAAACGAAAAACCTCTCTCAAA 1455  
QY 922 GCTGCGGCACTGAGGGCCCAAGGAAGTCTCTAGACCTGTGTGGAAGGCATTATGTGTGGA 981





QY 3109 GAAATTCCTCCAGTCCAGTCAATTAACAGAGTGGAGCATGCAGCTGAGTGAA 3168  
DB |||||  
QY 3676 AAAAAGTCTACCACTAGATTAATATGTAACAGACTGGAGTCAATGTCAGCTGAGTGAG 3735  
DB |||||  
QY 3169 AACGACCCCTGTGGTCAAGGCGTCAAGACCCGCTGTCTAAGCTGTGTGTGAGTATGGC 3228  
DB |||||  
QY 3736 AAGCGAGTTGTGGAAATGGAATAAAAAAGAGGATGTTGGATTGTGTTCCAAAGTATGGC 3795  
DB |||||  
QY 3229 AAGCCAGTCAGATGAGCAATGTGAGCAGCATTAATTTGGAGAGAGCCCGCAGAGATGAGC 3288  
DB |||||  
QY 3796 AAGTCAGTTGACTGAATATTTGGAAGCGTTGGCTTGGAGAAAGAACTGGCAGATGAAC 3855  
DB |||||  
QY 3289 ATTCCCTCTTGGTGAATGCGTGTCAACTGTCTCAGCTCTCAGGCTGGAAGCGCTTGGACA 3348  
DB |||||  
QY 3856 ACGTCTGATGCTGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGA 3915  
DB |||||  
QY 3349 GAGTGTTCACAGACCTGTGGCCATGAGGAGTGAATGAGCCGAGCTCGATTTATCATTTATG 3408  
DB |||||  
QY 3916 GATGTCTCTAAACATGTGGCTCTCAGGAAATATGATCCGAGACGAAACAGTGAACCCAG 3975  
DB |||||  
QY 3409 CCAACCCAGGAGAGAGGAGGCGCATGCCCCACAGAGCTTACCCAGGAGAGAAACCTGGCCA 3468  
DB |||||  
QY 3976 CCGTTTCAAGGTGATGGAAGACCATGCTCCCTGATGGACCATGTCACAAACCCCTGCCCA 4035  
DB |||||  
QY 3469 GTGACCCCTCTACAGCTGGTCTTGGCAACTGGTCTGATGATGAATGGAGGGTGA 3528  
DB |||||  
QY 4036 GTGAAGCCTTGTATCGGTGGCAATATGCGCAGTGGTCTCCATGCCAAGTGCAGGAGGCC 4095  
DB |||||  
QY 3529 GACTGTGGGAGGAGTTCAGATCCGACGCTTCTCTGATGGTCCACAGTGGTTCATTA 3588  
DB |||||  
QY 4096 CAGTGTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4155  
DB |||||  
QY 3589 TCTCATGAGCTGGACGCTGTGAGATGCACTGTGTGAGAGAAATGCCCTT----- 3638  
DB |||||  
QY 4156 GATGATTTTCAAGGAGTGGTGTGATGAGGATTTCTGTGCTGACATTTGATGATGATGAT 4215  
DB |||||  
QY 3639 -----TCAGGACAGCATCTGAGGAGCTGTGTTCTGTGCTTGCCTTGCAGGAGCTGCCAT 3693  
DB |||||  
QY 4216 GGTATTAATAATATGGTTCTGGAGAAATCTGTGACGACGCTTGCCTGAGTGTAT 4275  
DB |||||  
QY 3694 TTAACAGAAATGTCAGATGAGGAGCAGATGTAATTAACCTGATGATGAGGAGGAGGAGGAG 3753  
DB |||||  
QY 4276 TTGAAGGAGTGTCTTCTGAGGAGCTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 4335  
DB |||||  
QY 3754 GAGACTGTGGGCGCCGCTGAGTCAAGGAGCTTTTAAATTCAGTCTTTTGAAGACCA 3813  
DB |||||  
QY 4336 GCGTTTGTGGAATACAGGTCAAGTCCAGACCGGTGATTAACAGAACTAGAGATCAG 4395  
DB |||||  
QY 3814 GACAGCTGCCCCAACAGGTTCTAGAAACAGCCCTTGTACAGGAGGAGGAGGAGGAGGAGGAG 3873  
DB |||||  
QY 4396 CATCTGTGCCAGAGCAGATGTTAGAAACAAATCATGTTATGAGGAGGAGGAGGAGGAGGAG 4455  
DB |||||  
QY 3874 TACACATGGAAGAGAGTCTTTGGAGAAATACAGAAAGAACTGTATGTTGTCAGGCTTCA 3933  
DB |||||  
QY 4456 TATTAATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4515  
DB |||||  
QY 3934 GATGGCGTTTAAATGTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3993  
DB |||||  
QY 4516 GATGTATTAATGTAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4575  
DB |||||  
QY 3994 TGCATTTCAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4053  
DB |||||  
QY 4576 TGTAAACCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4635  
DB |||||  
QY 4054 GAGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4105  
DB |||||  
QY 4636 GAGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4695  
DB |||||  
QY 4106 -----TGAAAGTACAGGCTCAGAGGATAAAAAGCTGATGTGAAAGAACTTTCTG-G 4157  
DB |||||  
QY 4696 CCGGTGGTGGTATTACCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4755  
DB |||||

QY 4158 GAAAGACAGACCTGTGAATTCACAAATATACATGATATTTTAAAGGA--TGCTCTCTTCAA 4215  
DB |||||  
QY 4756 GTACATCCAAACCAACCTCCAGTAAACCCAGCAGGAGCGGAGGAGGAGGAGGAGGAGGAG 4815  
DB |||||  
QY 4216 CCATTTGATCCAGATGCGGAGTAAATTTGGGTTTATGGGTTTATGGGTTTATGGGTTTATGG 4275  
DB |||||  
QY 4816 CCATTTGGCCAGATGGGAGACTAAAGACCTGGGTTTACGGTGTAGCAGCTGGGCACTTT 4875  
DB |||||  
QY 4276 CTCATCATGATTTTCTAATATTTTACTTCTTACCTTGTGTTGCAAGAGCCA---AAAGCA 4332  
DB |||||  
QY 4876 GTGTACTACTCTTTTATTTGCTCTCCATGATTTATCTAGCTTGCATAAAGCAAGCAACC 4935  
DB |||||  
QY 4333 CATCAAGACACCTCCCAACAGAGCCTCTGACCTTAGCTACGATGAGAGACTTAGAC 4392  
DB |||||  
QY 4936 CAAGAGGAGCAAAACACGACTGAAACCTTTTACCTTAGCTATGATGAGAGATCCGAC 4995  
DB |||||  
QY 4393 ATGTAA 4398  
DB |||||  
QY 4996 ATGTAA 5001  
DB |||||  
RESULT 8  
ACCT2168  
ID ACC72168 standard; DNA; 6373 BP.  
XX ACC72168;  
AC ACC72168;  
XX 07-JUL-2003 (first entry)  
DT  
XX Human NOV47g coding sequence.  
DE  
XX Human; NOV; antidiabetic; anorectic; antibacterial; virucide;  
KW immunomodulator; cytostatic; nootropic; neuroprotective;  
KW antiparkinsonian; antipalemic; gene therapy; metabolic disorder;  
KW diabetes; obesity; infection; cachexia; cancer; dyslipidaemia;  
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
KW immune disorder; haematopoietic disorder; gene; ds.  
XX  
XX Homo sapiens.  
OS  
XX WO2003029423-A2.  
FN  
XX 10-APR-2003.  
PD  
XX  
PF 02-OCT-2002; 2002WO-US031358.  
XX  
XX 02-OCT-2001; 2001US-0325483P.  
PR 05-OCT-2001; 2001US-0327342P.  
PR 09-OCT-2001; 2001US-0327917P.  
PR 09-OCT-2001; 2001US-0328029P.  
PR 09-OCT-2001; 2001US-0328044P.  
PR 09-OCT-2001; 2001US-0328056P.  
PR 13-OCT-2001; 2001US-0328849P.  
PR 15-OCT-2001; 2001US-0329414P.  
PR 17-OCT-2001; 2001US-0330142P.  
PR 22-OCT-2001; 2001US-0341058P.  
PR 24-OCT-2001; 2001US-0339266P.  
PR 24-OCT-2001; 2001US-0343629P.  
PR 29-OCT-2001; 2001US-0349575P.  
PR 01-NOV-2001; 2001US-0346357P.  
PR 12-APR-2002; 2002US-0371972P.  
PR 12-APR-2002; 2002US-0371980P.  
PR 17-APR-2002; 2002US-0373261P.  
PR 19-APR-2002; 2002US-0373805P.  
PR 23-APR-2002; 2002US-0374738P.  
PR 15-MAY-2002; 2002US-0381101P.  
PR 17-MAY-2002; 2002US-0381635P.  
PR 23-MAY-2002; 2002US-0383830P.  
PR 01-OCT-2002; 2002US-00262839.  
XX  
XX (CURA-) CURAGEN CORP.  
XX  
XX Alsobrook JP, Anderson DW, Boldog FL, Burgess CE, Catterton E;  
PI





4396 CATCTGTGCCAGCAGCATGTTTGAACAAATAATCATGTTATGATGACAGTCTATGAA 4455  
3874 TACACATGGAAGCAAGTCTTTTGAACAATAAAGCAAGCACTGATGTCAGGCTTCA 3933  
4456 TATAATGATGCGCAGTCTTGGAGGGCTCTTCCGACAGTGTGGTCAAGGTCA 4515  
3934 GATGGCTTAATGTCACAGGAGCTGCTCCCTCAGGCGCTCTCTGCTGCCATTCGGCAG 3993  
4516 GATGGTATAAATGTAACAGGGGCTGTTGGTATGAGCAGCCTGATGCCAGAGTCT 4575  
3994 TGCAATCCAGCTGCGAGAAACCTTTCTCTACTGTACACAGGCTGAGTCTGTGTGT 4053  
4576 TGTAAACCCAGTGTAGTCAACCCACTGTCTGTAGCGAGACAAAACATGCCATGT 4635  
4054 GAGAAGGCTATACAGAGATAAATAAATCAATGTTTCTCTGATTAATGCA----- 4105  
4636 GAAGAAGGTTACACTGAAGTCATGTTCTTAACAGCACCTTGAAGCAATGCACATTATC 4695  
4106 -----TGAAGTACCAGCTCAGAGGATAAAGAGCTGATGTGAAAAACCTTCTG-G 4157  
4696 CCGTGGTGTATATACCCACTGAGGAGCAAAAGAGAGATGTGAAAACCAAGTGGCT 4755  
4158 GAAAAACAGACCTGTGAATCAAAAATACATGATATTTTAAAGGA--TGGTCTCTCAA 4215  
4756 GTACATCCAAACCAACCTCCAGTAACCCAGCAGGACGGGAGGACCTGTTTCTACAG 4815  
4216 CCACTTGATCCAGATGCGAGTAAATAATTTGGTGTATGCGGTTTCAGGTGGCGTTT 4275  
4816 CCAATTTGGGCGAGTGGGAGCTAAAGACCTGGGTTTACGGTGTAGCAGCTGGGCAAT 4875  
4276 CTCATCATGATTTCTCTAATATTTACTTCTACTGTTTGAAGAGCA--AAACCA 4332  
4876 GTGTTACTCATCTTATTTGTTCTCATGATTTATCTAGCTTGCAAAAGCAAGAACCC 4935  
4333 CATCAAGCACACTCCCAACAGAGCCTCTGACCTTAGCCTTAGCAGGAGACTTAGAC 4392  
4936 CAAAGAAGGCAAAACCAACCGACTGAAACCTTTAACTTAGCCTATGATGAGATCCGAC 4995  
4393 ATGTAA 4398  
4996 ATGTAA 5001

RESULT 9  
AI58451  
ID AAI58451 standard; cDNA; 6378 BP.  
CX AAI58451;  
CX 22-OCT-2001 (first entry)  
CX Human polynucleotide SEQ ID NO 654.  
CX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
CX peripheral nervous system; neuropathy; central nervous system; CNS;  
CX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
CX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
CX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
CX leukaemia; ss.  
CX Homo sapiens.  
CX WO200153312-A1.  
CX 26-JUL-2001.  
CX 26-DEC-2000; 2000WO-US034263.  
CX 23-DEC-1999; 99US-00471275.  
CX 21-JAN-2000; 2000US-00488725.  
CX 25-APR-2000; 2000US-00552317.  
CX 20-JUN-2000; 2000US-00598042.  
CX 19-JUL-2000; 2000US-00620312.

PR 03-AUG-2000; 2000US-00653450.  
PR 14-SEP-2000; 2000US-00662191.  
PR 19-OCT-2000; 2000US-00693036.  
PR 29-NOV-2000; 2000US-00727344.  
XX (HYSE-) HYSEQ INC.  
XX Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;  
XX Zhou P, Goodrich R, Drmanac RT;  
XX WPI; 2001-442253/47.  
XX P-PSDB; AAN39295.  
XX Novel nucleic acids and polypeptides, useful for treating disorders such  
XX as central nervous system injuries.  
XX Claim 1; SEQ ID NO 654; 10078pp; English.  
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the  
XX encoded polypeptides (AAM38642-AAM42213) with nootropic,  
XX immunosuppressant and cytostatic activity. The polynucleotides are useful  
XX in gene therapy. A composition containing a polypeptide or polynucleotide  
XX of the invention may be used to treat diseases of the peripheral nervous  
XX system, such as peripheral nervous injuries, peripheral neuropathy and  
XX localised neuropathies and central nervous system diseases, such as  
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
XX utilisation of the activities such as: Immune system suppression,  
XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening  
XX assays for receptor activity, arthritis and inflammation, leukaemia and  
XX C.N.S disorders. Note: The sequence data for this patent did not form  
XX part of the printed specification  
XX SQ Sequence 6378 BP; 1823 A; 1420 C; 1619 G; 1515 T; 0 U; 1 Other;  
Query Match 28.08; Score 1229.6; DB 4; Length 6378;  
Best Local Similarity 57.6%; Pred. NO. 0;  
Matches 2596; Conservative 0; Mismatches 1784; Indels 126; Gaps 17;  
QY 10 TGCATTGAGAGCTGAACCGAAGTGTGGTTGCAAAATATGCGAAGCACTTTTGCCTT 69  
DB 541 TGCATCCAGAAAGCAAGACATTCCTGCGAGGATATCATCTGTGAGTACTTTGAGCCC 600  
QY 70 CAGCTCTTACAGAACAGGCTTGGCTCATCTCTGTCCTCCCGGATTTGTAGTATCTGAG 129  
DB 601 AAGCTCTCTGAGCAGGCTTGGCTCATCTCTTGCAGCAAGATTGCAATCGTCTGAA 660  
QY 130 TTTCTTACATGTTCCAACTGTAGCAAGGATGTGGAGAAATTCAGCATAGAACTCGC 189  
DB 661 TTTTCTGCTGTTCCGATGCTCCAGACCTGCGGAGCGGCTCCAGCACCGGAGCGGT 720  
QY 190 GCGTCTATAGTCCCTCTCTTTGTTGTTGTTGCAATGTCCTCAAAATCTGAGTCAAGA 249  
DB 721 CATGTGTGGCGGCGCGAGTTCGGAGGCTCTGGTGTCTCAAACTCTGAGGAGTCCAG 780  
QY 250 GCCTGTGATGCTCCCTCTCTCTCTTGGGAGAGGAATATACATTTAGCCTTAAG 309  
DB 781 GTGTG-----CMAATCCAGTCCATGCGAGCGCGAGGCTCAGGTACAGCTGCAT 831  
QY 310 GTTGACCATGGAGTAAATGACAGCTGCTCATCTTTAAAGAAATTAATCAAGCGGAGA 369  
DB 832 GTGGGCGCTGGAGCAGCTGCTCAATGCCCACTCCCGACAGTAAGACAAGAGAGA 891  
QY 370 ACTGTTCTGGATTTTAA-----CTCTGATTTCAATGAG 402  
DB 892 CGCGGAGAGATTAAGAAACCGGAAAGGACCGCAGCAAGAGTAAAGATCCAGAGCC 951  
QY 403 CGAGTCACTTTTAAACATCAAGTTTACA-----AAGCATCATCATCTCGAAGTCT 453  
DB 952 CGCGAGCTTTTAAAGAAAGAGAAACAGAAACAGGACAGACAGCAAGACAAATAT 1011

454 TGGGAATAGAGATAGGTATCAAAACCGGAGGTTTCGTGTACAGAAAGTGTAGACAA 513  
1012 TGGGACATCCAGATTGGATATCAGACAGAGAGGTATGTGATTAACAAGACGGGAAA 1071  
514 AATGCTATGTTAAGCCTTTGGCTTCAAGATTCCCTCCATTGACTGTTCAGTCTCTGCATC 573  
1072 CTTGCTGATTTAAGCTTTTGGCCAGCAGAGAGCTTCCATGACTTCCAGTCTCTGTGTG 1131  
574 ATGCCCAAAGACTGTGAAACCTCCAGTGTCTCTGAGAGCCCTCTCTCCAAAGACATGC 633  
1132 ATCAACAAAGAGTCCAGGTTTCCGAGTGTGTGAGTGTGAGAGTGTGAGTGTGAGTGTG 1191  
634 CGTTCCAGGAGCTCTCTCCAGGATTTAGGAGCAGGAGCCGGAACGTGAGGACATGCTT 693  
1192 CATGACATGTTGCTCCCTGAGGACATCTGTGTAGGACACCAACCATCAGGAGTTTCCC 1251  
694 ATTGGAGTGAAGAGGAGTGTCTTGAATCTTGTGAGAAAGGCTCTGATTTGTGAAGGA 753  
1252 ATTGGCAGTGAAGAGGAGTGTCCAGATTTTGAAGAAAGAAACCTCTTGTCTCAAGGA 1311  
754 GAACTTCTGACGCAATGTCCAGGTATTTCTGAGAACTTCTGAGAACTTCTGAGAACTTCTG 813  
1312 GATGAGTGTGCTCCCTGTGCGACATGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1371  
814 GTCTCTCTCTCTCTCTGAGCAGCAGGATCCCACTGGCATGTGTGAGCAGCAGCAGTGTG 873  
1372 GTGGACCCCTTTGCTCAGTCAGCAGCAGCAAGAGGCGCGGCAACAGAGCGCCCTCTGTGGA 1431  
874 GGTGGATCCAGACCGGAGAGTGTACTGTGCCAG-----AGGTCACAGCA 921  
1432 GGGGACATCCAGACCGGAGAGTGTACTGTGCGTGTGAGGCCAACAGAAACCTCTCTCAAA 1491  
922 GCTGCGGCACTGAGGGCCCAAGAGTCTCTAGACTGTGTGAAAAGGCAATATGTGTGGGA 981  
1492 TTAAGTACCCACAGACAAAGAGCTCAAGCCAAATGAACTTAAATTTATGCACTGGA 1551  
982 CCGGCGGTTGCTCTCAGCTGTGCAATATCCCTGTCTCTAGGACTGTGATGATCT 1041  
1552 CCTATCCCTAATACTACACAGCTGTGCCACATCTCTGTGCACTGAACTGTGAGTGTCA 1611  
1042 TCTGTGTGAGCTGTGGGCTGTGATCCATGAAACTGTGATGAACTCTCAGGGGAAAAA 1101  
1612 CTTTGTGTGAGCTGTGGGACTGTGATTTATGAAACTGTGATGATGATGAGGAAAAA 1671  
1102 GGAATTTAGAACAGGACAGC-----CAATGTCTCTATGGAATCTACAGGCGCTGCA 1152  
1672 GGCCTTCAAACTGAGGAAGCGGCGATTAACCAATGAGCCACTGTGAGGCTCTGGGTTAAC 1731  
1153 GGGCATTTGCCCTCATTTGGTGGAGTGTCTTCTGTGAGGATCCATGTCTACCGATGG 1212  
1732 GGAATCTCCCTCACTTACTGGAAGCCATTCCTCTGGAAGAGCTGCTGTGTATGACTGG 1791  
1213 CTGGCA---TCAGAAAGGATCTGTTTCCCTGATCATGGA---AATGTGGCTGGGACAT 1266  
1792 AAAGCAGTGAGACTGGGAGACTGGAGCAGATTAACGAAAGAGTGTGTCCAGGACAG 1851  
1267 CGTATTTGAAGGCGCTGTGCAGATGACCGCGGAGAGTGTATCAGGAGTCTTTTC 1326  
1852 CAAGTTCAAGAGGTTGTGTGATCAACAGTGTATGAGAGAGTGTGACAGAGCTGTGC 1911  
1327 CCAGTTCCCTCTCTCTGAGAGGAGTCTGTGGAATTCCTGCGCAATGGAATGAGTGTG 1386  
1912 AGAGATGCCATCTTCCCATCTCTGTGGCTGTGATGCCCATGCGGAAAGACTGTGTG 1371  
1387 CTGAGCAGTGTGAGGAGTGTGATCTGTGTTCCAGTCTGTGTTCAATAAATACTCAGAT 1446  
1972 CTCAGCAGATGTTACGTGTCTCTCTGTCTCACACACCTGCTCAGGGAAAAAGACAGAA 2031  
1447 GGGAAACAGACAGGTCAAGAACTATCTGTGC-----ACTGGCTGGGAGTGTGAAAGCCA 1503  
2032 GGGAAACAGATACAGGACAGATCAATCTGTGCGCTATGCGGCTGAGAGGTTGAAATTCG 2091  
1504 TGTCCCTCTAGTCAGGCTCTCCAGAGCATCGTTTGTGTAAATGACCATTCCTGTATGCA 1563

2092 TGTCAAAATAGCAGTGTCTTTCAGAGAGTCAAGAGCTGTAAATGAGCATCTTGTGCACAGTG 2151  
1564 CTTTCACTGGAGACATCGCTTGGGGCCCTTGTCTGAGGACACATTGTTAACTGCCCTT 1623  
2152 TACCACTGGCAACTGGTTCCTTGGGGCCAGTGCATTGAGGACACCTCAGTATCGTCTTC 2211  
1624 AATCAACCATTTGGCTGGAATGGAGAGCCAGTGTGTGTAGGATTCAGACTCGGAGA 1683  
2212 AACCAACTTACGACTTGGAAATGGGAGGCTCTCTGTCTGTGCGCATGCAGACAAGAAA 2271  
1684 GTCCTCTGTGTCAAGAGTCACTGGGACAAAGTAAATGACCAAAAGATGCCAGATTCCTACT 1743  
2272 GTCACTGTGTGCGAGTCAATGTGGCCAGAGTGGGACCCAAAATGTCTCTGAAAGCCTT 2331  
1744 CGACTCTGAACTGTGCGCCCTCTGTTTCTCCCATGCAAAAAAGACTGTATTTGTGACTGCT 1803  
2332 CGACTCTGAACTGTAAAGCCTTGTCTGCTTCTTGTAAAGAGACTGTATTTGTGACCCCA 2391  
1804 TTCAGTGTGTGACACCTTGGCCCAAGGATGTCCNAGCAGGAATGCCACAGT---AAAA 1860  
2392 TATAGTGTGTGACATCATGCCCCCTCTCTGTGTAAAGAGGGGACTCCAGTATCAGGAAG 2451  
1861 CAGTCTCTGATACAGAAATCATCATCAAGAGCAGCAATGGAGGCCAGAAATGCCAGAT 1920  
2452 CAGTCTAGGATCGGCTCATCTTCACTGCTGCCAGCCACGGGGCCGAGACTGCACAGAT 2511  
1921 ACCTTATATGAGCAGAGAGTGTGAAGATGTTTCTTGTGTCTGTATATCTGTTGAG 1980  
2512 CCCCTCTATGAAGAGAGGCTGTGAGGACCTCAAGCGTCCAAAGCTACAGGTGGAAG 2571  
1981 CCACAGAAATGAGGCCCTTGCACTTGTAGTGCAGAGTCTGTCTGCGAGGAATAACGGGC 2040  
2572 ACTCACAAATGGCGCAGATGCCAATTAAGTCTCTGAGCGTGCACACAGACAGCCCTGA 2631  
2041 AGCAGTGAAGCCTGTGGAAGAGGTTCAAAACAGAGCTGTCTCATGCTCTCTGATGAC 2100  
2632 GCACAGGAAGGCTGTGGGCTGGGCGCAGCAGCAGAGCCATTACTTGTGCGAAAGCAAGT 2691  
2101 AACCGCTCAGCAGAAATGATGAATGCTCAAGCAGACAAACGGCATGCTCTCTCTGTG 2160  
2692 GGAGACAGGCTGGAAATCCATGAGTGTCTCAGTATGACAGCCCTGTGCCAGCCCTTACC 2751  
2161 CAAGATGCAAGTCCCATGTGTGAGAGACTGCACCTTCACTGTCTGTGTTGTTCAAGTTACG 2220  
2752 CAGGCTGTCCAGATCCCTGCGCAGGATGACTGTCAATGACAGCTGTGTTCAAGTTTCT 2811  
2221 CCTGCTCCAGACTGTGAAGCCACAAAAGTAGGCGGCGACAGCTCACAGGGAAGC 2280  
2812 TCATGCAATGGAGACTGTGTGAGTGTAGGACCCAGAAAGCGCACTCTTGTGGAAAAAGT 2871  
2281 AGAAAGAAAGGAGAAATGCCAGGATTTCTGACCTTTACCTCTAGTGGAGACAGAACTATGT 2340  
2872 AAAAAAAGGAGAAATGTAAGAAATTTCCCATTTGTATCCCTGATTGAGACTCAGTATGT 2931  
2341 CTTGTGTGATGATTTATATCCCACTTATGGAACCTGTGAGAACTGTGATGATGCTTCTCCAGAA 2400  
2932 CTTGTGCAATATATTAATGCAACCTGTGGGAACTGCTGAGACTGTATTTTACCAGAG 2991  
2401 GGCAAGAGGAGCCCTCACCGAGGACTGGGCTGCAAGCAGACAGCAAGAAATGTGGAGAA 2460  
2992 GGAAGTGGAGTGTCTGCTGGAAATGAAGTACAGGAGACATCAAGAAATCGGACAA 3051  
2461 GGCCTGCGTTTTCAGAGAGTGTGCTGTGATAAAAATGGAAGACCTGTGTGACCCCTCC 2520  
3052 GGATATCGTTTACCAAGCAATGGCATGTACGATCAAAATGGCAGCTTGTGGAAAACTCT 3111  
2521 TTTCTCAGCAGCTGTGTTTACATCAAGAAAAATGTGTCATTCTCCCTGCCATTTGATGTC 2580  
3112 AGATTAACAGCCATGGTTTACATTGAGAGGCTGTGATCATCCCTGCGCTCAGACTGC 3171  
2581 AAGTTAAGCGATGTGTCTGATTTGGGGTCTTTCAGATTCATCTTGTGGAATTTGGAGTGA 2640



3172 AAGCTCAGTGGAGTGGTCCAACTGGTGGCTGAGCAAGTCTGTGGAGTGGTGTGAAG 3231  
2641 ATTGATCCAAATGGCTAAAGAAACCTTCAATGGAGGAGGACCATGTCCCAACTG 2700  
3232 GTTCGTTCTAAATGGCTGGCTGAAACACCATATAATGGAGGAGGCTTGGCCCAACTG 3291  
2701 GATC-----TCAAGAAATCAGGTACATGAGGAGTCCCATGTTTACAGTGAAGTCAATCAG 2754  
3292 GACCATGTCAACAGGACAGAGTGTATGAGTGTGCCATGCCACAGTCACTGCAACACAG 3351  
2755 TATTCCTGGTGTAGAACACTGCTTCAATGCAAAATCAAAATGAGTGAAGTCCCTG 2814  
3352 TACCTATGGGTACAGAGCCCTGGAGCATCTGCAAGGTGACCTTGTGAATATGCGGAG 3411  
2815 CGCTGTGGAGGAGGAAACAATCTAGAAATACAGATGTGT---GAATACCTGGATGT 2871  
3412 AACTGTGGAGGGCTGAAACCCGAAAGTGAATGATGCATGCAGATACAGAGATGCC 3471  
2872 GAAGTGGAGCATGGATAGCAACCTGTGCAACAGGATGAATTCCTCCAGAAACCCAG 2931  
3472 CCTTCTGAACATGTAGAGGATTAACCTCTGTGACCCAGAAAGAGATGCCCTGGGCTTGA 3531  
2932 TCCTGTTCTCTATGTCCCAATGAGTGTCTCATCTGAGTGGGACTTTGGAGCAAA 2991  
3532 GTGTGCAATACCATGCCCTGAGGACTGTGTGATATCTGAATGGGTCAATGGACCCAA 3591  
2992 TGCCACAGTCATGCCATCCCAACAAT---GCAGAGAACTCGCCACCTGTCTAAGA 3048  
3592 TGTGTTTGGCTTGGCAATCAAGCAGTTTCCGGCAAAAGGTGAGTCCATCCCATCAGACAA 3651  
3049 CCATCACTGAACCTCAGGACTGTGTGAAGCTCACAGTGGCGCTTGCCTCTGTGAT 3108  
3652 CCAGCTGATGAAGAGATCTTGCCTTAATGCTGTTGAGAAAGAACCTGTGAACCTGAAC 3711  
3109 GAAATATGCTTCAGATTCAGATCAATCTAAACAGAGTGGAGCACATGCCAGCTGAGTGAA 3168  
3712 AAAAACTGCTACCATATGATTAATGTAACAGACTGGAGTACATGTGAGTGTGAGT 3771  
3169 AAGCAGCCCTGTGGTCAAGCGCTCAGACCCGCTGTGATGCTGTGTGTCAGTGTATGCC 3228  
3772 AAGCGATGTTGTGGAATGGAATAAAAAAGAGATGTTGGATGTTGTCGAAAGTGTATGCC 3831  
3229 AAGCCAGTCAGCATGGACCAATGTGAGCAGCATAAATTTGGAGAGCCCGCAGAGATGAGC 3288  
3832 AAGTCAGTTGACCTGAATATTGTGAAGCGCTTGGCTTGGAGAGAACTGGCAGATGAAC 3891  
3289 ATTCCCTGCTGTGGAAATGCGTGGTCAACTGTGACCTCTCAGGTGGAACCGCTTGGACA 3348  
3892 ACGTCTCTGCATGTGGAAATGCCCTGTGAACCTGTGACCTTCTGATTTGGTCTCCTTGGTCA 3951  
3349 GAGTGTTCACAGACCTGTGGCCATGGAGTTCGAATGAGCGGAGTCAATTTATCATTTAG 3408  
3952 GAATGTTCTCAACATGTGGCTTCAAGGAAATGATCCGAGAGCAACAGTGAACCCAG 4011  
3409 CCAACCCAGAGAGAGGAGCGCCATGCCACAGAGCTTACCCAGGAGAAACCTGCCCA 3468  
4012 CCCTTTCAAGGTGATGAGAGACCATGCGCTTCCCTGATGGACAGTCCAAACCTGCCCA 4071  
3469 GTGACCCCTCTACAGCTGGTCTTGGCAATGTGCTGCTGATGTAATTTGGAGGTGGA 3528  
4072 GTGAAGCTTGTATCGGTGGCAATATGGCCAGTGGTCTCCATGCCAAGTGCAGGAGGCC 4131  
3529 GACTGTGGGGAAGAGTTCAGATCCGAGCGCTTCTCTGATGTGCTCCACAGTGGTTCATA 3588  
4132 CAGTGTGAGAGAGGACAGNACAGAGAAATTTCTGTGTAGTGAATGATGGTCACT 4191  
3589 TCTCATGCACTGACGTGTGAGGATGCACTGTGTGGAGAAATGCCCCCTT----- 3638  
4192 GATGATTTCAAGAAAGTGGTGAATGAGGAAATCTGTGCTGACATTTCAATATAGAT 4251  
3639 -----TCAGGACAGCATCCTGGAACAGCTGTGTTCTGTGCTTGGCCAGGAGACTGCCAT 3693  
4252 GGTAATAAAAATATGGTCTTGGAGGAATCTCTGAGCCAGCCTTGGCCAGGTGACTGTTAT 4311

QY 3694 TTAACAGATGGTCCAGAGTGGAGCACATGTGAATTAACCTGCATGTTGATGGAAGACTTT 3753  
Db 4312 TTGAAGACTGGTCTTCTCGAGCTGTGACCTGACCTGTGTAATGTTGAGGATCTA 4371  
QY 3754 GAGACTGTGGCCGCCAGTCTAGATCAAGGACTTTTATTAATCAGTCTTTTGAGAACAA 3813  
Db 4372 GGCTTTGGTGAATACAGGTCCAGATCCAGACCGGTGATTATACAAGAACTAGAGAACTAG 4431  
QY 3814 GACAGCTGCCCCCAACAGTCTTAGAAACACAGCCCTTGTACAGGAGCAAAATGTTATCAC 3873  
Db 4432 CATCTGTGCCCAGAGCAGATGTTAGAAACAAATCATGTTATGATGACAGTGTCTATGAA 4491  
QY 3874 TACACATGGAAGCAAGTCTTTTGGAAACAATAACGAAACAACTGTATGTCGTCACAGTTC 3933  
Db 4492 TATAATGGATGGCCAGTGTCTGGAAGGCTCTTCCCGAACAGTGTGTGTCRAAGGTCA 4551  
QY 3934 GATGGCTTAATGTACAGAGAGGTGTCTCCCTCAGGCCGTCTCTGTCGATCGCATGCGGAG 3993  
Db 4552 GATGGTATAAATGTAAACAGGGGGCTGCTTGGTGAATGAGCCAGCCTGATGCCGACAGTCT 4611  
QY 3994 TGCATTCCAGCTTCAGAAAAACCTTTCTCTACTGTACACAGGGTGGAGTCTGTGGTGT 4053  
Db 4612 TGTAAACCCAGCTGTACTCAACCCACTCGTACTGTAGCGAGACAAAAACATGCCATTGT 4671  
QY 4054 GAGAAGGCTATACAGAGATAAATGAATCAAAATGAAATGAAATGAAATGAAATGAAATGAA 4105  
Db 4672 GAAGAAGGCTACACTGAAGTCAATGCTCTTAAACAGCACCCTTGAGCAATGCACACTTATC 4731  
QY 4106 -----TGAAGTACAGGCTCAGAGGATAAAAAAGCTGATGTGAAAAACCTTCTG-G 4157  
Db 4732 CCGCTGGTGGTATTACCAACATGAGAGACAAAGAGAGATGTGAAAAACAGTCGGGCT 4791  
QY 4158 GAAAAACAGACCTGTGAATTCAAAAATACATGATATTTTAAAGGA--TGGTCTCTTCAA 4215  
Db 4792 GTACATCCAAACCCCAACCTCCAGTAAACCCAGCAGCGGAGAGGACCTGCTTTCTACAG 4851  
QY 4216 CCATTGATCCAGATGGCCGAGTAAATTTGGGTTTATGGCTTTTACAGTGGCGCTTTT 4275  
Db 4852 CCAATTGGCCAGATGGAGACTAAAGACCTGGGTTTACCGTGTAGCAGCTGGGGCATTT 4911  
QY 4276 CTCATCATGATTTCTCTAAATTTACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 4332  
Db 4912 GTGTACTCATCTTTATGCTCTCCATGATTTATCTAGCTTGCAGAAAGCCAAAGAACCC 4971  
QY 4333 CATCAAGACACCTCCCAACAGAACCTCTGACCTTAGCTTACGATGAGACTTAGAC 4392  
Db 4972 CAAAGAGGCAAAACCAACCGACTGAAACCTTTAACTTAGCTTATGATGAGATGCCGAC 5031  
QY 4393 ATGTAA 4398  
Db 5032 ATGTAA 5037

RESULT 10  
ADB48422  
ID ADB48422 standard; cDNA; 6378 BP.  
XX AC ADB48422;  
XX DT 04-DEC-2003 (first entry)  
XX DE Novel human cDNA SEQ ID NO 332.  
XX KW ss; cancer; neurodegenerative disease; human.  
XX OS Homo sapiens.  
XX PN US2003104529-A1.  
XX PD 05-JUN-2003.  
XX PF 04-JAN-2002; 2002US-00037270.

XX	21-JAN-2000; 2000US-00488725.	DB	1072	GCTGCTGATTAAAGCTTTTGCCAGCAAGAGAGCTTCCAATGACCTTCCAGTCTCTGTGTG	1131
PR	25-APR-2000; 2000US-00552317.	QY	574	ATGCCCAAGAGCTGTGAAACCTCCAGTGGTCTCTCGAGCCCTGCTGCAAGACATGC	633
PR	19-JUL-2000; 2000US-00620312.	DB	1132	ATCACCRAAGAGTCCAGGTTTCCGAGTGGTTCAGAGTGGAGCCCTGCTCAAAACATGC	1191
XX	(ZHOU/) ZHOU P.	QY	634	CGTTCAGGAGTCTCTTCCAGAGATTAGAGAGAGGAGCCGGAACGTAAGACATGGCT	693
PA	(TANG/) TANG Y T.	DB	1192	CATGACATGGTGTCCCTGCGAGGACCTGCTGTAAAGGACACGAACCATCAGGACGTTTCCC	1251
PA	(LIUC/) LIU C.	QY	694	ATTGAGAGTGAAGAGAGTGTCTGNACTTCTCAGAAAGAGGCTGCTCATTTGTGAAGGA	753
PA	(ASUNDI V.) ASUNDI V.	DB	1252	ATTGACAGTGAAGAGAGTGTCCAGAAATTTGAAGAAAGAACCTGTTTGTCTCAGGA	1311
PA	(DRMA/) DRMANAC R T.	QY	754	GAATTTCTGCAGCAATGTCCAGGTATTCCTGGAGAACTTCTGAATGGAAAGATGCCAA	813
XX	Zhou P, Tang YT, Liu C, Asundi V, Drmanac RT;	DB	1312	GATGAGTGTGCTCCCTGTGCCACGTATGGCTGGAGAACTACAGAGTGGACTGAGTGGCGT	1371
XX	WP1; 2003-678194/64.	QY	814	GTCTCTCTCTCTCGAGCAGCAGGATCCCACTGGCATGTGACGGGACCCGTGTGTGGC	873
XX	New polynucleotide, useful for treating diseases e.g., cancer or	DB	1372	GTGGACCTTTGCTCAGTCAGCAGGACAAGAGGCGGGCAACACGAGCGCCCTCTGTGGA	1431
XX	neurodegenerative diseases.	QY	874	GGTGGGATCCAGACCCCGGAGGTGTACTGTGCCAG-----AGCGTACCAGCA	921
XX	Claim 1; SEQ ID NO 332; 99pp; English.	DB	1432	GGGGCATCCAGACCCCGAGAGTGTACTGCGTGCAGGCCAACGAAACCTCTCTCACA	1491
XX	The invention relates to a polynucleotide comprising a sequence given in	QY	922	GCTCGGCACCTGAGGGCCACGAAAGTCTCTAGACCTGTGGAAAGGCATTTATGTGGGA	981
XX	the specification, or its mature protein-coding portion, or its	DB	1492	TTAAGTACCCCAAGAAACAAAGAGCCTCAAGCAATGGACTTTAAATATGACATGGA	1551
XX	complement. The polynucleotide is useful for treating diseases e.g.,	QY	982	CCCGCCCGCTTGGCCCTCTCAGCTCTGCAATATCCCTTGTCTACGGACTGCATAGTATCT	1041
XX	cancer or neurodegenerative diseases and many others listed in the	DB	1552	CCATATCCCTAATACTACACAGCTGTGCCACATTCCTTGTCCAATGATGTGAAGTTCA	1611
XX	specification. The present sequence represents a novel human cDNA. Note:	QY	1042	TCCTGTGAGCCTCGGGCCTGTGCATCCATGAAACTGTGTATGAACTCAGGGGAAAAA	1101
XX	The sequence data for this patent did not form part of the printed	DB	1612	CCTTGTGAGCTTGGGACCTTGTACTTATGAAACTGTATGATCAGCAGGAGAAAAA	1671
XX	specification but was obtained in electronic format directly from USPTO	QY	1102	GGATTGAAACGAGCGACG-----CCATGTCTCATGGAATCTACAGGCGCTGCA	1152
XX	at seqdata.uspto.gov/sequence.html?docID=20030104529.	DB	1672	GGCTTCAACTGAGAGAGCGGCGCATTAACAAAGCCACTGGAGGCTCTGGGGTAACC	1731
XX	Query Match 28.0%; Score 1229.6; DB 8; Length 6378;	QY	1153	GGGCAATCCCTCAATTTGGTGGAGTCTGTTCCTTGTGAGGATCCAATGTCTACCGATGG	1212
XX	Best Local Similarity 57.6%; Pred. No. 0;	DB	1732	GGAAACTGCCCTCACTTACTGGAAGCCATTCCTGTGGAAGAGCCTGCTGTATGACTGG	1791
XX	Matches 2596; Conservative 0; Mismatches 1794; Indels 126; Gaps 17;	QY	1213	CTGGCA---TCAGAGAGGATCTGTTTCCCTGATCATGGAA---NATGTGGCTGGGACAT	1266
QY	10 TGCATTGAGAGCTGAACCGAAGTGTGGTGGCAATGAATATGCGAACTTTGGCCCTT	DB	1792	AAAGCAGTGAAGACTGGGAGACTGGAGCCAGATAACCGAAAGGAGTGTGGTCCAGGCAG	1851
DB	541 TGCATCAGAAAGACAAAGACATTCCTGCGAGAGATATCATCTGTGAGTACTTTGAGCC	QY	1267	CGTATTCTGAAGGCGCTCTGCAGAAATGACCGCGAGAAAGATGTATCAGGAGTCTTTTC	1326
QY	70 CAGCCTCTCAGAACAGGCTGCTCATTCCTTGTCCCGGATGTTGTAGTATCTGAG	DB	1852	CAAGTTCAAGAGGTTGTGTGCATCAACAGTGTGAGAGAGAGTTGACAGACAGCTGTGC	1911
DB	601 AAGCCTCTCCTGGAGAGGCTTGCTCTCTTCTTCCCTGAGCAAGATTGCATCGTGTCTGAA	QY	1327	CGAGTTCCCTCTCTCTGAGAGGAGTCTTGTGAAATTCCTCCCGAATGGAGTGTGTG	1386
QY	130 TTCTTACCATGTCCTCACTGTAGCAAGGAGTGTGGGAGAAATTTGACGATAGAACTCGC	DB	1912	AGAGATGCCATCTTCCCCTGCTGTGCTGTATGCCCATGCCCATGCCGAAAGACTGTGTG	1971
DB	661 TTTTCTGCTGTCGGAATGCTCAAGACCTGCGGACGCGGCTCCAGCACCGGACGCT	QY	1387	CTGAGCAGTGGAGCGGAGTGTGTCTCTGTTCCAGTCTCTTCAAAATAAAACTCAGAT	1446
QY	190 GCGGTGATGCTCCCAATTTCTGCTCTTGGGGAAGAGGAATATACATTTAGCCTTAAG	DB	1972	CTCAGCAGATGGTCTTACGTGTCTCTCTGCTCAACACCTGCTCAGGGAAACACAGAA	2031
DB	781 GTGTG-----CCAATCCAGTCCATGCGAGCGCGAGGAGCTCAGGTACAGCTGCAT	QY	1447	GGGAAACAGACCAAGTCAAGAACTATCTGCG-----ACTGGCTGGGGAAGTGGAAAGCCA	1503
QY	310 GTTGACCATGGAGTAATGAGACTCCCTCATCTTTAAAGAAATTAATCAAGCGGAAGA	DB	2032	GGGAAACAGATACAGACAGCATCCATTCGCGCTATGCGGCTGAAGAGGTGGAATTCGC	2091
DB	832 GTGGGGCCCTGGAGACCTGCTCAATGCCCACTCCCGACAAGTAAGACAAGCAAGGAGA	QY	1504	TGTCCCTCTGCTCAGGCTCTCCAGAGCATCGTTTGTGTATGACCATTCCTGTATGCAG	1563
QY	370 ACTGTTCTGGAATTTAA-----CTCTGATTCAAATGAG	DB	2092	TGTCCTCAATAGAGTGTCTTTCAGAAAGTAGCAAGCTGTATGAGCATCTTTGACAGTG	2151
DB	892 CGCGGGAAGATAAGAACCGGAAAGAGACCGGACGAGAGTAAGGATCCAGAGCC	QY	1564	CTTCACTGGGAGACATCGCCTTGGGGCCCTTGTCTGAGGACATTTGGTAACCTGCCCTT	1623
QY	403 CGAGTACCTTTAAACATCAAGTTACA-----AAGCATCATCATTCGAAAGTCT	DB	2152	TACCACTGGCAAACTGGTCCCTCGGGCCAGTGCATTTGAGGACACCTCAGTATCGTCTTC	2211
DB	952 CGCGAGCTTTAAGAAAAGAGAAAACAGAAAACAGGACAGACAGACAAGAGACAATAT	QY			
QY	454 TGGCAATAGAGTAGTATCAACCCGCGAGTTCGTGTACAGAAAGTGTAGCAAA	DB			
DB	1012 TGGACATCCAGATTGATATCAGCCAGAGAGGTATGTGCATTTACAAAGACGGGAAA	QY			
QY	514 AATGCTATGTTAAGCCTTTGCTTTCAAGATTCTCTCCATTTGACTGTTTCTGAGTCT	DB			

1624 AATGCAACCATTTGGCTGGAAATGGAGAAAGCCACGCTGGTGTAGGCAATTCAGACTCGGAGA 1683  
1625 AATGCAACCATTTGGCTGGAAATGGAGAAAGCCACGCTGGTGTAGGCAATTCAGACTCGGAGA 1684  
2212 AACACAACATACGACTTTGGAAATGGGAGGCTCTGCTGTCTGGCAATGCAGACAAGAAAA 2271  
1684 GTCTTCTGTGTCAAGAGTCAAGTGGGACAAAGTAATGACCAAAAGATGTCCAGATTCTACT 1743  
2272 GTCACTGTGTGGAGTCAATGTGGGCAAGTGGGACCCAAAAATGTCTCTGAAGGCTT 2331  
1744 CGACCTGAACTGTGGGCGCTCTGTTTCTCCATGCAAAAGACTGATTTGTGACTGCT 1803  
2332 CGACCTGAACTGTGAGGCTCTGTTCTCTCTGTTGTAAGAGGACTGTATTGTGACCCCA 2391  
1804 TTCAGTGTGACACCCCTGCCCCAAGGATGTCGACGAGGAAATGCCACAGT---AAAA 1860  
2392 TATAGTGTGACATCATGCCCCCTCTCTGTGTAAGAGGAGCTCCAGTATCAGGAAG 2451  
1861 CAGTCTCGATACAGAAATCATCCAAAGAGCAGCCAAATGGAGGCCAGGAATGCCAGAT 1920  
2452 CAGTCTAGGCATCGGCTCATCTCAGTGCAGCCCAACCGGGGCGAGACTGCACAGAT 2511  
1921 ACCTTATATGAGGAGAGAGTGTGAGATGTTCTCTGTGTCCTGTATATCGGTGGAG 1980  
2512 CCCCCTATGAAGAGAGGCTGTGAGGCACCTCAAGCGTGCCAAAGCTTACAGGTGGAAG 2571  
1981 CCACAGAAATGGAGCCCTTGCATCTTAGTGCCAGAGTCTGTCTGCGAGGAAATACCGGC 2040  
2572 ACTCAAAATGGCGCAGATGCCAAATTTAGTCCCTTGGAGCGTGCACAGACAGCCCTGGA 2631  
2041 AGCAGTGAAGCTGTGGAAGGGTTTAAACAGAGCTGTCTCATGATCTCTGATGAC 2100  
2632 GCACAGGAAGGCTGTGGGCTCGGCGACAGGCAAGGCAATTAATTTGCGCAAGCAAGAT 2691  
2101 AACCGGTGACGAAATGATGAATGCCCTCAAGCAGACAAAGCGCATGCTCTCTCTGTG 2160  
2692 GGGAGACGGCTGGAATCATGAGTGCCTACAGTATGCAAGGCGCTGTGCGAGCCCTTACC 2751  
2161 CAAGATGACAGTCCCATGTGCGAAGACTGCACCTTCACTGTGTGTCGAAAGTTTACG 2220  
2752 CAGGCTGCGAGATCCCTGCGAGGATGACTGCAATTGACAGCTGGTCCAAGTTTCT 2811  
2221 CCTGTCTCACGAACTGTGAAAGCCCAAAAGTAGGCGGCGACAGCTCACAGGGAAGC 2280  
2812 TCATGCAATGGAGACTGTGGTGCAGTTAGGACCAAGAAAGCGCACTCTTGTGGAAGAAAGT 2871  
2281 AGAAGAGGAGAAATGCCAGGATCTGACCTTTACCTCTAGTGGAGACAGAACTATGT 2340  
2872 AAAAAAAAAAAAAAAAAATGTAATAATTCCTATTTGTATCCCTGATTTGAGACTCAGTATTGT 2931  
2341 CTTGTGATGAATTTATATCCCAACCTTATGGAAGACTGGTCAAGTTCATTTCCAGAA 2400  
2932 CTTTGTGACAAATATAATGCACAACCTGTGGGAACTGGTCAGACTGTATTTTACCAAG 2991  
2401 GGCAGAGGAGGCTCACCGAGACTGGGGTACAGAGCAGACAGCAAGAAATGTGGAGAA 2460  
2992 GGAAGTGGAGTGTGTGCGGAATGAAGTACAGGAGCATCAAGGAATGGGACAA 3051  
2461 GGCCTGCGCTTTCGAGCACTACCTGTCTGATTAATAATGGAAGACCTGTTGAOCCTTCC 2520  
3052 GGATATCGTTACCAAGCAATGGCAGTACGATCAAAATGCGAGCTGTGTGGAACATCT 3111  
2521 TTCTGAGCAGCTCTGTTTACATTCAGAAATAATGTGTCATTTCCCTGCCATTTGATGTC 2580  
3112 AGATGTAACAGCAATGGTTTACATTTAGGAGGCGCTGCATCATCCCCCTGCCCTCAGACTGC 3171  
2581 AAGTTAAGCAATTTGTTAGTTGGGGTCTTTCAGTTTCTTGTGGAATTTGAGTGTAGA 2640  
3172 AAGCTCAGTGTGAGTGTGCAACTGGTGGGCTGCGGCTGCAGCAAGTCTGTGGAGTGTGAAG 3231  
2641 ATTGATCCAAATGCTAAAGAAAAACCTTACATGAGGACGACCATGTGCCAAACTG 2700  
3232 GTTCGTTCTAAATGGCTGCGTGAATAAACATATAATGAGGAGGAGGCTTGTGCCCAAACTG 3291

2701 GATC-----TCAGAAATCAGGTACATGAGGCAAGTCCCATGTTTACAGTGTAGTCAATCAG 2754  
3292 GACCATGTCAACAGGCACAGGTGTATAGGTTTCTCCATGCCACAGTGTGACTGTCAACCCAG 3351  
2755 TATTCCTGGGTGTAGAACTACTGCTCTCATGCAAAATCAAAATCAGCTGAGCTGAGGTCCCTG 2814  
3352 TACCTATGGTTCACAGAGCCCTGGAGCATCTGCAAGGTGACCTTTGTGAATATCGGGAG 3411  
2815 CGCTGTGAGGAGAAACAAATCTAGGAAATCAGATGTGT---GAATACTGCGGATGGT 2871  
3412 AACTGTGAGAGGCGGTGCAACCCGAAAGTGAATGATGATGAGAAATACAGCAGATGGC 3471  
2872 GAAGTGGAGCAGTGGATAGCAACCTGTGCAACAGAGTGAATTTCCCCAGAAAACCCAG 2931  
3472 CTTTCTGAACATGTAGAGATTACCTCTGTGACCCAGAGAGATGCCCTTGGGCTCTAGA 3531  
2932 TCCTGTTCTCTTATGTCTCCCAATGAGTGTGTCTGATCTGAGTGGGACTTTTGGAGCAAA 2991  
3532 GTGTGCAAAATTACCATGCCCTGAGGACTGTGTGATATCTGATGGGTCCTCATGACCCAA 3591  
2992 TGCCCAAGTCAATGATCCCAACAAAT---CGAGAGAAAGAACTCGCCACCTGTGAAG 3048  
3592 TGTGTTTGTGCTTGCATCAAGCAGTTCGCGCAAGGTCAGCTGATCCCATCAGACAA 3651  
3049 CCATCACTGAACCTCAAGGACTGTGTGAAAGACTCAAGGTGAGCTTGCCTCTCTGAAT 3108  
3652 CCAGCTGATGAAGAGATCTTGCCTTATGCTGTTGAGAAAGAACCTGTAACTGAAC 3711  
3109 GAAAAATGCTTCCAGTTCAGTACAATCTAAACAGTGTGAGCAATGCGCAGCTGAGTGAA 3168  
3712 AAAAACTGTCTACCACTATGATTAATGTAACAGACTGGAGTACATGTGAGCTGAGTGAG 3771  
3169 AACGACCTGTGTCAAGCGCTCAGGACCCGCTCTAAGCTGTGTGTGAGTGTGAGTGTGAG 3228  
3772 AAGCAGTTTGTGAAATGGAATPAAAAAAGAGATGTTGATTTGTTTGAAGTGTGAGTGTG 3831  
3229 AAGCCAGTCAAGTGAACCAATGTGAGCAGCAATAATTTGAGAGAGCCCGCAGAGAAATGAG 3288  
3832 AAGTCAGTTGACCTGAAATATTTGTGAAGCGCTTGGCTTGGAGAAAGAACTGGCAGATGAAC 3891  
3289 ATTCCCTGCTTGTGGAATGCGTGTGCTAACTGTGAGCTCTCAGGTTGGAGCGGTGAGCA 3348  
3892 AGTCTGTGATGTGGAATGCCCTGTGAACTGTGAGCTTCTGATTTGCTCTCTGCTGCTG 3951  
3349 GAGTGTTCACAGACCTGTGTGCCATGAGGCTGCAATGAGCGGACTCGATTTATCATATG 3408  
3952 GAATGTCTCAACATGTGGCTTCAAGGAAATAATGATCCGAGAGCAAGACAGTGAACCCAG 4011  
3409 CCAACCCAGGAGAGGACGCGCATGCCCAAGAGCTTACCAGGAGAAACCTGCCCA 3468  
4012 CTTTCAAGGTGATGGAAGACCATGCCCTTCCCTGATGGAACCAAGTCCAAACCTGCCCA 4071  
3469 GTGACCCCTGTCTACAGCTGGGTCTTGGCACTGTGTCATGTATAATTTGAGGAGTGA 3528  
4072 GTGAGCTTGTATCGTGGCAATGATGCCAGTGTCTCCATGCCAGTGTGAGGAGGCC 4131  
3529 GACTGTGGGAGAGGATCAGATCCGACGCTTCTGTGATGTCATGTCACAGTGTGTTCAATA 3588  
4132 CAGTGTGAGAGAGGAGCCAGAAACAGGAAACATTTCTGTGTAGTAAGTGTGATGAGTGTG 4191  
3589 TCTCATGACGCTGAGCTGTGCGAGGATGCACTGTGTGGAAGAAATGCCCTT----- 3638  
4192 GATGATTTAGCAAAATGGTGGATGAGGAATTTCTGTGCTGACATTTGAATCTCATATAGAT 4251  
3639 -----TCAGGACAGCATCTGAGCAGCTGTGTCTGTGCTTGTGCCCTTGGCCAGGAGACTGCCAT 3693  
4252 GGTAAATAAAAAATATGTCTTGGAGGAATCCTGCGAGCCAGCTTGGCCAGTGTGAT 4311  
3694 TTAAACAGATGTCAGAGTGGAGCAGATGTGAATTAACCTGCAATGATGGAAGAGACTTT 3753  
4312 TTGAAGAGACTGTGTTTCCCTGGAGCCCTGTGTGAGTGTGATGAGTGTGAGGATCTTA 4371  
3754 GAGACTGTGGGCGCGCAGTCTAGATCAAGGACTTTTATAATTCAGTCTTTTGAAGAACCA 3813

```
Db 4372 GCCTTGGTGAATACAGGTCAGATCCAGACCGGTGATTTATACAGAACTAGAGATCAG 4431
2Y 3814 GACAGCTGCCCCCAACAGGTTCTTAGAAACACGCCCTTGTACAGAGGCAAAATGTTATCAC 3873
Db 4432 CATCTGTGCCAGACAGATGTTTAGAAACAAAATCATGTTATGATGACAGTGTCTATGAA 4491
2Y 3874 TACACATGGAAGCAAGCTCTTTGGACAAATAACGAAGCACTATGATGTCAGCCTTCA 3933
Db 4492 TATTAATGGATGGCCAGTGTGTTGAAGGGCTCTCCGAAACAGTGTGGTCAAGGTCA 4551
QY 3934 GATGGCGTTAATGTACAGAGGCTGCTCCCTCAGCCCGTCTCTGTCGCAATGCGCAG 3993
Db 4552 GATGGTATAAATGTAAACAGGGGCTGCTTGTGATGAGCCAGCCTGATGCCAGAGTCT 4611
QY 3994 TGCATTCAGCCTCGAAGAACCTTTCTCTACTGTATACACAGGCTGGAGTCTGTGGTCT 4053
Db 4612 TGTAAACCCCGTGTAGTCAACCCCTCGTACTGTAGCGAGACAAACATGCCATTGT 4671
QY 4054 GAGAAGGGCTATACAGAGATAAATGAAATCAAAATGCTTCTCGATTACTGCA----- 4105
Db 4672 GAAGAAGGTACACTGAAGTCACTGCTTCTTAACAGCACCTTGAGCAATGCACACTATC 4731
QY 4106 -----TGAAGTACAGGCTCAGAGATAAAGAGCTGATGTGAAAACCTTTCTG-G 4157
Db 4732 CCGGTGTGTATTATCCACCATGGAGGACAAAGAGAGATGTGAAAACCTGCGGCT 4791
QY 4158 GAAAAACAGACCTGTGAATTCAAAAATACATGATATTTTAAAGGA--TGGTCTCTTCAA 4215
Db 4792 GTACATCCAACCCCAACCTCCAGTAAACCCAGCAGGAGGAGGAGGAGGAGGAGGAG 4851
QY 4216 CCATTTGATCAGATGCGCGAGTAAATAATTTGGTTTATGGGTTTACAGTGCGGTTTT 4275
Db 4852 CCATTTGGGCGAGATGGAGACTAAAGACCTGGTTTACGGTGTAGCAGTGGGGCATTT 4911
QY 4276 CTCATCATGATTTTCTTAATTTTACTTCTACCTGTTTGCAGAGGCA---AAACCA 4332
Db 4912 GTGTTACTCATCTTTATGTTCTCCATGATTTATCTAGCTTGCAAAAGCCAAAGAACCC 4971
QY 4333 CATCAAGCACACCTCCCAACAGAGCCTCTGACCTTAGCCTACGATGAGACTTAGAC 4392
Db 4972 CAAGAAGGCAAAACACCGACTGAACCTTTAACTTTAGCCTATGATGATGAGATCGGAC 5031
QY 4393 ATGTAA 4398
Db 5032 ATGTAA 5037

RESULT 11
AAI60237
AC AAI60237 standard; cDNA; 6292 BP.
AC AAI60237;
XX 22-OCT-2001 (first entry)
XX Human polynucleotide SEQ ID NO 4226.
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia; ss.
XX OS Homo sapiens.
XX PN WO200153312-A1.
XX XX 26-JUL-2001.
XX PF 26-DEC-2000; 2000WO-US034263.
XX XX
```

```
PR 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00498725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX (HYSE-) HYSQ INC.
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX WPI: 2001-442253/47.
DR P-PSDB; AAM41081.
XX Novel nucleic acids and polypeptides, useful for treating disorders such
PT as central nervous system injuries.
PT Claim 1; SEQ ID NO 4226; 10078pp; English.
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
XX SQ Sequence 6292 BP; 1818 A; 1391 C; 1595 G; 1487 T; 0 U; 1 Other;
Query Match 27.1%; Score 1190.4; DB 4; Length 6292;
Best Local Similarity 57.5%; Pred. No. 0;
Matches 2594; Conservative 0; Mismatches 1786; Indels 132; Gaps 20;
QY 10 TGCATTCAGAGCTGAACCGAACTGTGTTGCAATGAAATATCGGAACACTTTGCCCTT 69
Db 510 TGCATCCAGAAAGACAAAGACATTCCTCGGAGGATATCATCTGTGAGTACTTTGAGCCC 569
QY 70 CAGCCTCCTACAGACAGGCTTGCTCATTCCTTGTCCCGGATTTGTAGTATCTGAG 129
Db 570 AAGCCTCTCCTGGAGCAGGCTTGCTCATTCCTTGTCCAGCAAGATTGCTGTCTGAA 629
QY 130 TTCTTTACCATTGTCACACTGTAGCAAGGGATGTGGGAAGAAATTCGAGCATAGAACTCGC 189
Db 630 TTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 689
QY 190 GGGGTATAGCTCCCTCTCTTTGGTGGTTGCAATGTCCAAATCTGACTGAGTACAGA 249
Db 690 CATGTGGTGGGCGCCCCCGCAGTTTCGAGGCTCTGCTGTGTCAAAACCTTGACGGATTCAG 749
QY 250 GCGTGTGATGTCCTCCATTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 309
Db 750 GTGTG-----CCATCCAGTCCATGCGAGCGCGAGAGCTCAGGTACAGCTGCAT 800
QY 310 GTTGACCATGAGTAATGAGACTGCTGCTCATCTTAAAGAAATTAATCCAGCGGAGA 369
Db 801 GTGGGGCCCTGGAGCACCTGCTCAATGCCCTCCCTCCAGCAAGTAAGCAAGCAAGGAGA 860
QY 370 ACTGTTCTGGATTTTAA-----CTCTGATTCAAATGAG 402
Db 861 CCGGGGAAGATAAAGAACGGGAAAGACCGGACCAAGAGTAAGAGTAAAGGATCCAGAGCC 920
```

403 CGAGTCACTTTAAACATCAAAAGTTACA-----AAGCACATCATATTCGAAGTCT 453  
1b  
921 CGGAGCTTATTAAAGAAAGAGAAACAGAAACAGCGACAGACAGAGAACAAATAT 980  
1b  
454 TGGGCATAGATAGATAGTTATCAACCCGGAGGTTCTGTGTACAGAGATGATGGACAA 513  
1b  
981 TGGGACATCCAGATTGGATATCAGACCCAGAGAGGTTATGTGATTTAACAAAGCGGGAAA 1040  
1b  
514 AATGCTATGTTAAGCCTTTGCGCTTCAAGATTCTTCCCATTCAGTGTTCAGTCTGCGATC 573  
1b  
1041 GCTGCTGATTTAAGCTTTTGGCCAGAGAGAGCTTCCNATGACCTTCCAGTCTGTGTG 1100  
1b  
574 ATGCCCAAGATCTGTAACCTCCAGTGGTCTCTCTGGAGCCCTGTCTCAAAGACATGC 633  
1b  
1101 ATCACCAAGAGTGCAGGTTTCCGAGTGGTCAGAGTGGAGCCCTGTCTCAAAAACATGC 1160  
1b  
634 CGTTCAGGAGTCTCTTGGCCAGGATTTAGAGCAGGAGCGGAACGTGAAGCACATGGCT 693  
1b  
1161 CATGACATGGTCTCCCTCGAGCACTCGTGTAAAGACAGAACCATCAGGCGATTTCCC 1220  
1b  
694 ATTGGAGTGGAAAGAGTGTCTGAACTTCTTGAGAAAGAGGCTGTGATTTGAAGGA 753  
1b  
1221 ATTGGCAGTGAAGAGGTTCCAGAAATTGAAGAAAGAAAGAACCTGTGTCTCAAGGA 1280  
1b  
754 GAACTCTGCACCAATGTCCAGGATTTCTCGAGAACTTCTGAAATGGAAGATGCCAA 813  
1b  
1281 GATGGAGTGTCCCTGTGCCAGTGTGCTGGAGAACTACAGAGTGGACTGAGTGCCGT 1340  
1b  
814 GTCTCTCTCTCTCTCGAGCAGGAGTATCCCATGTGGCATGTGACGGGACCCGTGTGGC 873  
1b  
1341 GTGGACCTTTGCTCAGTCAGCAGGACAGAGCGCGGCAACAGACGCGCTCTGTGGA 1400  
1b  
874 GGTGGATCAGACCCGGAGGTTACTGTGCCAG-----AGCTACAGCA 921  
1b  
1401 GGGGGCATCCAGACCCGAGAGTGTACTCGTGCAGCGGCAACGAAACCTCTCTCAAA 1460  
1b  
922 GCTGCCGACTCAGGCGCAAGGAAGTCTCTAGACCTGTGGAAGGCAATTATGTGGGA 981  
1b  
1461 TTAAGTACCCACAGAACAAAGAGCTCAAGCCNATGGACTTAATATGCACTGGA 1520  
1b  
982 CCGGCCCTGTGCTCAGCTGTGCAATATCCCTTGCTCTACGGAGTGCATATATCT 1041  
1b  
1521 CTTATCCCTAATACTACACAGCTGTGCCACATCTCTGTGCCAACTGAATGTGAATTTCA 1580  
1b  
1042 TCTGTCTCAGCTGGGCGCTGTGCATCCATGAATACTGTCATGACCTCAGGGAAGAAA 1101  
1b  
1581 CTTGTGAGCTTGGGACCTTGTACTTAAGAACTGTAATGATCAGCAAGGGAAGAAA 1640  
1b  
1102 GGATTTAGAACGAGGCGAGCG-----CCATGTCCTCATGGAATCTACAGGCGCTGCA 1152  
1b  
1641 GGCTTCAAACTGAGGAAGCGGCGCATTTACCAATGAGCCCACTGGAGGCTCTGGGTAACC 1700  
1b  
1153 GGGCATGCGCTCATTTGGTGGAGTCTGTCTCTGTGAGGATCCATGTGCTACCGATGG 1212  
1b  
1701 GGAATGCGCCCTCACTTACTTGAAGCCATTCCTGTGTAAGAGCCCTGCTGTTATGACTGG 1760  
1b  
1213 CTGGCA---TCGAAGGAGTCTGTTTCCCTGATCATGGA---AATGTGGCTGGGACAT 1266  
1b  
1761 AAAGCAGTGAGACTGGGAGCTGGAGCCAGATACGGAAGAGAGTGTGTCAGGACG 1820  
1b  
1267 CGTATTTGAAGCCGTCTGCGAAGATGACCGGAGAGAGATGTATCAGGAGTCTTTGC 1326  
1b  
1821 CAAGTTCAAGAGGTTGTGTGCATCAACAGTGTGGAAGAGAGTGTGACACAGACGTGTGC 1880  
1b  
1327 CAGATTTCCCTCTCTCAGAGGAGTCTGTGAAATTTCCCTGCGGATGGAAGTGTGTG 1386  
1b  
1881 AGAGATGCCATTTTCCCATTCCTGTGGCTGTGTATGTCGCCCCATGCCGAAGACTGTGTG 1940  
1b  
1387 CTGAGCGAGTGAAGGAGTGTGCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1446  
1b  
1941 CTGAGCACATGCTGTACGTGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2000

1447 GGGAAACAGACCAAGTCAAGAACTATCTCTGCC---ACTGCTGGGGAAGGTGGAAGCCA 1503  
1b  
2001 GGGAAACAGATACAGACACGATCCATCTCTGCTGCTATCGGGTGAAGAGTGAATTCG 2060  
1b  
1504 TGTCCCTCTAGTCAGGCTCTCAAGAGCATCTGTTGTGTATGACCATCTCTGTATGCGAG 1563  
1b  
2061 TGTCCAAATAGCATGCTTTTGCAGAGAGTACGAAGCTGTAAATGAGCATCTCTTGCACAGTG 2120  
1b  
1564 CTTCACTGGGAGACATCGCTTGGGGCCCTTGTCTTGAGGACACATTTGTAATGCGCCTT 1623  
1b  
2121 TACCACCTGGCAACTGGTCTCTGGGCCAGTGCATTTAGGACACCTCAGTATCGTCTTC 2180  
1b  
1624 AATGCAACCATTTGGCTGGAATGGAGAGCCACCTGTGTGTGTAGGCACTCAGTATCGAGT 1683  
1b  
2181 AACACAACTACGATTTGGAATGGGGAGGCTCTCTGCTGTGTGGGATGCGAGACAAGAAA 2240  
1b  
1684 GTCTTCTGTCTCAAGAGTCAAGTGGGACAGTAATGACCAAAAGATGTCCAGATTTCTACT 1743  
1b  
2241 GTCATCTGTGTGCGAGTCAATGTGGGCCAAGTGGGACCCMAAANAATGTCTGAAAGCCTT 2300  
1b  
1744 CGACCTGAAACTGTGGGCCCTCTTCTTCTCCATGCAAAAAGACCTGTATTGTGACTGCT 1803  
1b  
2301 CGACCTGAAACTGTAAAGCCTTGTCTGCTTCTTGTAAAGAGACTGTATTGTGACCCCA 2360  
1b  
1804 TTCAGTGAAGTGGACACCTGCCCAAGGAT--GTGCCAAGCAGGAATGCCACAGT---AAA 1859  
1b  
2361 TATAGTACTGGACATCATGCCCTCTTATCGTGTAAAGAGGGGATCTCCAGTATCAGGAA 2420  
1b  
1860 ACAGTCTCGATACAGAACTCATCCAAAGAGCAGCAATGGAGGCGCAGGAATGCCCAGA 1919  
1b  
2421 GCAGTCTAGGCATCGGTCTCATCTTTCAGCTGCCAGCAACGGGGGCGGAGACTGCCAGA 2480  
1b  
1920 TACCTTATATGAGAGAGAGAGTGTGAAGATGTTCTGTTGTCTGTATATCGGTGG-- 1977  
1b  
2481 TCCCTCTATTAAGAGAGAGGCTGTGAGGCACCTCAAGCGTGCCTCAAGAGTACAGGTGGGA 2540  
1b  
1978 -AAGCCACAGAAATGGAGCCCTTGGCATCTTAGTGCC--AGAGTCTGTCTGGCAGGGAATA 2034  
1b  
2541 AGACTCAAAATGGGCACAGATGCCAAATTAGTCTCTTTGGAGCGTGCACACAGACAGCC 2600  
1b  
2035 ACGGGACAGAGTGAAGCCTGTGAAAGGGGTTCACAAAGAGCTGTCTCATGCACTCTCT 2094  
1b  
2601 CTTGGAGCACAGGAAGGCTGTGGGCTGTGGGCGCAGGCAAGAGCCATTACTTGTGCGAAG 2660  
1b  
2095 GATGACAAACCGGTACAGAGAAATGATGGAATGCTCAAGCAGACAAACGGCATGCTCTC 2154  
1b  
2661 CAAGTGGAGGACAGGCTGGAATCCATGAGTGTCTACAGTATGAGGCGCTGTGCCAGCC 2720  
1b  
2155 CTTGTGCAAGAATGCAAGTCCCATGTTCGAGAAAGACTGCACCTTCTACTGTCTGGTCCAAG 2214  
1b  
2721 CTTTACCCAGGCTGCCAGATCCCTCTCCAGGATGACTGTCTCAATTGACCAAGCTGCTCCAAG 2780  
1b  
2215 TTTACGCCCTGCTCCAGAACTGTGAAGCCACAAAAGTAGGCGGCGACAGTCAAGGG 2274  
1b  
2781 TTTTCTTTCATGCAATGAGAGCTGTGTGCAAGTGTAGGACCAAGAAAGCGCACTCTTGTGGA 2840  
1b  
2275 AAAAGCAAGAAAGAGAGAAATGCCAGGATTTCTGACCTTTTACCCTCTAGTGGAGACAGAA 2334  
1b  
2841 AAAAGTAAAGAGAGAAATGTAAATTTCCATTTGTTATCTCCCTGATTGAGACTCAG 2900  
1b  
2335 CTATGCTCTTGTGATGATTTATATCCAAACCTTATGGAACCTGGGTCAAGTTGCATCTTT 2394  
1b  
2901 TATTGTCTTGTGCAAAATATAATGACCAACCTGTGGGAACTGGGTCAAGCTGTATTTTA 2960  
1b  
2395 CCAGAAGGCGAAGAGGAGCCCTCACCGAGGACTCGGCTTCAAGCAGACAGCAAGAAATGT 2454  
1b  
2961 CCAGAAGGAAAGTGAAGTGTGTCTGGGAATGAAGTACAGAGACATCAAGGAATGC 3020  
1b  
2455 GGAGAAGCCCTGCGCTTTTCAGCAGTAGCTGTCTCTGATATAAATGAAAGACCTGTGTGAC 2514  
1b  
3021 GGACAGGATATCGTTTACCAGCAATGGCATGTCTACGATCAAAATGGCAGGCTTGTGGAA 3080  
1b  
2515 CCCTCTTCTGACAGCAGCTCTGTTGTATCTCAAGAAANAATGTGTCTATTCCTGCCCCATTT 2574

Db 3081 ACATCTAGATGTAACAGCCATGGTTACATTTAGAGAGCCCTGCATCATCCCTCCCTCA 3140  
Qy 2575 GATTGCAAGTTAAGCGATTGGTCTAGTTGGGGTCTTGGCAGTTTCATCTTGTGAAATTGGA 2634  
Db 3141 GACTGCAAGCTCAGTGAAGTGGTCAACTGGTGGCGTGGAGAGTCTTGTGGAGTGGT 3200  
Qy 2635 GTGAGAAATTGATCCAAATGGCTTAAAGAAACCTTTACAAATGGAGGACCAATGTCCC 2694  
Db 3201 GTGAAGGTTGGTTCTAAATGGCTGGCTGAAACCATATATATGAGAGAGGCGCTTGCC 3260  
Qy 2695 AAACCTGGATC-----TCAGAATCAGTATCATGAGGAGTCCCATGTATACATGAGTGC 2748  
Db 3261 AAACCTGCAATGTCACAGGAGGAGTATGAGGTTGTCCCATGCCACAGTACTGC 3320  
Qy 2749 AATCAGTATTCTGGGTTGTAGAACACTGGTCTTCATGCAAAATCAACAATGAGCTGAGG 2808  
Db 3321 AACCACTACCTATGGGTCACAGAGCCCTGGAGCATCTGCAAGGTGACCTTTGTGAATATG 3380  
Qy 2809 TCCCTGGCTGTGGAGGAGGACACATCTTAGGAATCAGTGTGT-----GAATCTGG 2865  
Db 3381 CGGAGAACTGTGGAGAGGCGTCAAAACCCGAAAGTGAATGATGCAAGAAATCAGCA 3440  
Qy 2866 GATGGTGAAGGTGAGCAGTGGATAGCAACCTGTGCAACAGGATGAAATTTCCCCAGAA 2925  
Db 3441 GATGCCCTCTGAACATGTAGAGGATTACCTCTGTGACCCAGAGAGATGCCCTGGGC 3500  
Qy 2926 ACCAGTCTCTTCTTATGTGTCCCAATGAGTGTGTCTATGTCTGAGTGGGACCTTTGG 2985  
Db 3501 TCTAGAGTGTCAAAATTAACCATGCCCTGGAGACTGTGTGATATCTGAATGGGTCATG 3560  
Qy 2986 AGCAAAATGCCACAGTCAATGGATGCCCAACACAAATGACAGAGAG-----AATCTGCCACCTG 3042  
Db 3561 ACCCAATGTGTTTGGCTTGCATCAAGACAGTTTCCGGCAAGGTGAGTGTATCCCATC 3620  
Qy 3043 CTAAGACATCACTGAATCAAGACTGTGTGCAAGACTGTGCAAGACTCAAGTGTGAGCTTGCCT 3102  
Db 3621 AGACAACAGCTGATGAAGAGAGATCTTGCCCTAAATGCTGTGAGAAAGAACCTGTAA 3680  
Qy 3103 CTGAATGAAATTCCTTCCAGTTCCTGATCAATCAACAGTGGAGACATGCCAGCTG 3162  
Db 3681 CTGAACAAACTCTACCTACTATGATTAATGTAACAGACTGGAGTACATGTACAGCTG 3740  
Qy 3163 AGTGAAACCGACCTGTGTCTCAAGGCGTCAGGACCGCTCTGCTTAAGCTGTGTGTCAGT 3222  
Db 3741 AGTGAGAAGGCAGTTTGTGGAATGGAATAAAACAAGGATTTGGATTGTGTTGGAAGT 3800  
Qy 3223 GATGGCAGCCAGTCAGATGGACCAATGTGACAGCATTAATTTGGAGAGCCCCAGAGA 3282  
Db 3801 GATGGCAAGTCAGTTGACCTGAAATATTTGGAAGCGCTTGGCTTGGAGAGAACTGGCAG 3860  
Qy 3283 ATGAGCATTCCTGCTTGGTGGAAATGGCTGTCAACTGTGAGCTCTCAGGGTGGAGCGCT 3342  
Db 3861 ATGAACAGCTCTGCAATGGTGAATGCCCTGTGAACTGTGAGCTTCTGATTGCTCTCT 3920  
Qy 3343 TGGACAGAGTGTTCACAGACCTGTGGCCATGGAGGTGGAATGAGCCGGACTCGATTATTC 3402  
Db 3921 TGGTCAGAATGTTCTCAAACTGTGGCTTCACAGGAAATAATGATCCGAAGCAACAGT 3980  
Qy 3403 ATTATGCCAACCCAGAGAGAGGAGCGGCATGCCCCACAGAGCTTACCCAGGAGAAAC 3462  
Db 3981 ACCCAGCCCTTCAAGGTGATGGAAGACCATGCCCTTCCCTGATGGACCACTCCAAACC 4040  
Qy 3463 TGCCAGTGAACCCCTGCTCAGCTGGGTCTTGGCAACTGGCTGATGATGAAATGGAG 3522  
Db 4041 TGCCAGTGAAGCCCTTGTATCGGTGGCAATATGGCCAGTGTCTCCATGCCAAGTGCAG 4100  
Qy 3523 GGTGAGACTGTGGGAGGAGTTCAGATCCGACCTTTCCTGATGCTGCCACAGTGT 3582  
Db 4101 GAGGCCAGTGTGGAGAGGAGGACAGCAAAAGGAACATTTCTTGTGTAGTGAATGGG 4160  
Qy 3583 TCAATATCTCATGACGTGGACGTGTGAGGATGCATGTGTGAGAAATGCCCTT----- 3638

Db 4161 TCAGCTGATGATTTTCAGCAAAAGTGGTGGATGAGGAATTTCTGTGTGACATTTGAACCTATT 4220  
Qy 3639 -----TCAGGACAGCATCTGAAGCAGCTGTGTCTGTGCTTGGCCAGGAGAC 3687  
Db 4221 ATAGATGGTAATAAAATATGTTCTGGAGAAATCTCGACGACCACTTGGCCAGGTGAC 4280  
Qy 3688 TGCCATTTAAACAGAAATGCTGAGATGGAGCAGATGTGAATTAACCTGATTTGATGGAGA 3747  
Db 4281 TGTATTATTGAAGGAGTGTCTTCTGGAGCGCTGTGTGAGCTGACCTGTGTGAATGGTGA 4340  
Qy 3748 AGCTTTGAGACTGTGGGCGCCAGTCTAGATCAAGGACTTTTATAATTCAGTCTTTTTCAG 3807  
Db 4341 GATCTAGCTTTGGTGGATATACAGTCAAGTCCAGACCGGTGATATACAGAACTAGAG 4400  
Qy 3808 AACCAAGACAGCTGCCCAACAGAGTTCTTGAACAACGCCCTTGTACAGGAGGCAATGT 3867  
Db 4401 AATCAGCATCTGTGCCAGAGCAGATGTTGAACAACAAATCATGTATGATGAGCAGTGC 4460  
Qy 3868 TATCACTACATGGAAGCAAGTCTTTGGAAACAATAACGAACCACTGTATGTGTCAG 3927  
Db 4461 TATGATATTAATGATGGCCAGTCTTTGGAGGGCTCTTCCCAACAGTGTGTGTGTCAA 4520  
Qy 3928 CTTTCAGATGCGCTTAATGTTCAGAGAGGCTGCTCCCTCAGGCCCTCTCTGTGCCATT 3987  
Db 4521 AGTTCAGATGTTATAAATGTAACAGGGGCTGCTTGGTATGAGCCAGCTGATGCCGAC 4580  
Qy 3988 CGGAGTGCATTCAGCCTGACAGAAACCTTTCTCTACTGTACACAGGGTGGAGTCTGT 4047  
Db 4581 AGTCTTTGTAACCAACCGTGTAGTCAACCCCACTGCTACTGTAGCGAGACAAAACATGC 4640  
Qy 4048 GCTTGTGAAGGGCTTATACAGAGATTAATGAAATCAAAATGTTTCTCTGGATTACTGCA-- 4105  
Db 4641 CATTTGTGAAGAGGTACACTGAAGTCACTGCTTCTTAACAGCACCTTGAGCAATGCACA 4700  
Qy 4106 -----TGAAGTACCAGGCTCAGAGATTAATAAGATGATGTGAAAAACCTT 4152  
Db 4701 CTTATCCCGCTGGTGGTATTACCCACCATGGAGGACAAAAGAGGAGATGTGAAAAAC 4760  
Qy 4153 TCTG--GGAAAAACAGACCTGTGAAATTCAAAAATACATGATATTTTAAAGGA--TGGTCT 4209  
Db 4761 CGGGCTGTGATCCAAACCAACCTCCAGTAAACCCAGCAGGACGGGAGACCTGGTTT 4820  
Qy 4210 CTTCAACCACTGTATCCAGATGGCGAGTAAAAATTTGGGTTTATGGGTTTACAGTGGC 4269  
Db 4821 CTACAGCCTTTGGCCAGATGGGAGACTAAAGACCTGGGTTTACGGTGTAGCAGCTGG 4880  
Qy 4270 GCTTTTCTCATCATGATTTTCTTAATTTTACTTCTTCTTCTTCTTCTTCTTCTTCTTCT 4326  
Db 4881 GATTTGTGTTTACTCATCTTTTATTTGTCCTCATGATTTTATCTAGCTTGCATAAAGCCAA 4940  
Qy 4327 AAACCAATCAAGCAGACCTCCCAACAGAGGCTCTGACCTTTAGCCTTACGATGGAGAC 4386  
Db 4941 AAACCCCAAGAGGCAAAAACAAACCGACTGAAACCTTTAACCTTAGCCTATGATGGAGAT 5000  
Qy 4387 TTAGACATGTAA 4398  
Db 5001 GCCGACATGTAA 5012

RESULT 12  
ACCT2167  
ID ACCT2167 standard; DNA; 4647 BP.  
XX AC ACCT2167;  
XX DT 07-JUL-2003 (first entry)  
XX Human NOV47f coding sequence.  
XX Human; NOV; antidiabetic; anorectic; antibacterial; virucide;  
KW immunomodulator; cytosstatic; nootropic; neuroprotective;  
KW antiparkinsonian; antilipemic; gene therapy; metabolic disorder;  
KW diabetes; obesity; infection; cachexia; cancer; dyslipidaemia;



neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
immune disorder; haematopoietic disorder; gene; ds.

Homo sapiens.

W02003029423-A2.

10-APR-2003.

02-OCT-2002; 2002WO-US031358.

02-OCT-2001; 2001US-0326483P.

05-OCT-2001; 2001US-0327342P.

09-OCT-2001; 2001US-0327917P.

09-OCT-2001; 2001US-0328029P.

09-OCT-2001; 2001US-0328044P.

09-OCT-2001; 2001US-0328056P.

12-OCT-2001; 2001US-0328849P.

15-OCT-2001; 2001US-0329414P.

17-OCT-2001; 2001US-0330142P.

22-OCT-2001; 2001US-0341058P.

24-OCT-2001; 2001US-0339286P.

24-OCT-2001; 2001US-0343629P.

29-OCT-2001; 2001US-0349575P.

01-NOV-2001; 2001US-0346357P.

12-APR-2002; 2002US-0371972P.

12-APR-2002; 2002US-0371980P.

17-APR-2002; 2002US-0373261P.

19-APR-2002; 2002US-0373805P.

23-APR-2002; 2002US-0374738P.

16-MAY-2002; 2002US-0381101P.

17-MAY-2002; 2002US-0381635P.

29-MAY-2002; 2002US-0383830P.

01-OCT-2002; 2002US-00262839.

(CURA-) CURAGEN CORP.

Alsobrook JP, Anderson DW, Boldog FL, Burgess CE, Catterton E;

Edinger SR, Ellerman K, Gerlach Vn, Gorman L, Guo X, Ji W;

Kekuda R, Leach MD, Li L, Miller CE, Patturajan M, Rieger DK;

Rothenberg ME, Shimketa RA, Smithson G, Spytek KA, Taupier RJ;

Vernet CAM, Voss EZ, Zernhusen BD, Zhong M;

WPI; 2003-381625/36.

P-PSDS; ABR58455.

NOVX polypeptides and nucleic acids useful for diagnosing, preventing or

treating NOVX-associated disorders, e.g. diabetes, obesity, cancer or

dyslipidemia, and in chromosome mapping, tissue typing or

pharmacogenomics.

Claim 20; Page 259-260; 487pp; English.

The present invention relates to novel human NOV proteins and their

coding sequences (ACC72075-ACC72181 and ABR58469). The NOV

proteins are useful in manufacturing a medicament for treating a syndrome associated with a human disease. The NOV proteins and coding sequences may be used to diagnose, treat or prevent metabolic disorders such as diabetes or obesity, infections, cachexia, cancer, neurodegenerative disorders such as Alzheimer's disease or Parkinson's disease, immune disorders, haematopoietic disorders and various dyslipidaemias

Sequence 4647 BP; 1312 A; 1072 C; 1279 G; 984 T; 0 U; 0 Other;

Query Match 26.8%; Score 1177.4; DB 7; Length 4647;

Best Local Similarity 57.3%; Pred. No. 0;

Matches 2439; Conservative 0; Mismatches 1711; Indels 105; Gaps 13;

2y 10 TGCATTTCAGAGCTGACCCGAACTGCTGGTGAATGAATATGCGAACACTTTGCCCTT 69

Db 277 TGCATCAGAAAGCAAGAACATTCCTCGCGAGGATATCATCTGTGAGTACTTTGAGCCC 336

2y 70 CAGCCTCCTACAGAACAGGCTTGCTTCATTCCTTGTCGCCGGATTGTGTAGTACTGTAG 129

Db	1408	GGCTTCAAAC	TGAGGAAGCGGCGCATTA	CCAAATGAGCCCACTGGAGGCTCTGGGGTAA	CC	1467
Qy	1153	GGGCATTGCCCT	CATTTGGTGGAGTCTGT	TTCCTTGTGAGGATCCAAATGTGCTACCGATGG	1212	
Db	1468	GGAAACTGCCCT	CACCTTACTTGAAGCCAT	TCCCTGTGAAGAGCTGCCGTGTATGACTGG	1527	
Qy	1213	CTGGCA---	TCAGNAGGGATCTGTTT	CCCTGTATCATGGAA--	AATGTGGCCTGGGACAT	1266
Db	1528	AAAGCAGTGA	CTGGGAAATCGGACCCAGAT	TAACGGAAAGAGTGTGGTCCAGGCACG	1587	
Qy	1267	CGTATTCTGA	AGCCGCTCTGCCAAGATGA	CCGGGAGGAAGTGTATCAGGAGTCTTTGG	1326	
Db	1588	CAAGTTCAAG	AGGTTGTGTGCATCA	ACAGTGAATGGAGAGAGTTGACAGACAGCTGTGC	1647	
Qy	1327	CCAGTTC	CCCTCCTGAGAGGAAGTCT	TGTGAATTCCTCTGCCGAATGAGACTGTGTG	1386	
Db	1648	AGAGATGCCA	CTTCTCCCATCCCTGT	GGCCTGTGATGCCCGCTGCCGAAGACTGTGTG	1707	
Qy	1387	CTGAGCAGT	GGGAGTGGTCACTCTGT	TCCAGTCTGTTCAAAATAAAACTCAGAT	1446	
Db	1708	CTCAGCACAT	TGGTCTAGCTGGTCTCT	GTCTCACACCTGTCTCAGGGAATAACGACAGAA	1767	
Qy	1447	GGGAAACAG	ACCAGGTCAAGAACTAT	CCTGGC---	ACTGGCTGGGAGAGTGGAAAGCCA	1503
Db	1768	GGGAAAACAG	ATACGAGCACGATCCA	TCTGGCCTATGCGGGTGAAGAGTGGAAATTCG	1827	
Qy	1504	TGTC	CCCTAGTCAGGCTCTCC	AGAGCATGCTTGTGTAAAGACCATTCCTGTATGCAG	1563	
Db	1828	TGTC	CCAAATAGCAGTGTCT	TTCGAGAAGTACGAAGCTGTAAAGAGCATCCTTCGACAGTG	1887	
Qy	1564	CTTCACT	GGGAGACATCCCTTGGG	CCCTGTCTTGAGGACACATTTGGTAACTGCGCCTT	1623	
Db	1888	TACCACT	TGGCAAACTGGTCT	CCGCGCAGTGCATTGAGGACACCTCAGTATCGTCTTC	1947	
Qy	1624	AATGC	ACCAATTGGCTGGAA	TGGAAGCAACGTCGTGGTGTAGGCATTCAGACTCGGAGA	1683	
Db	1948	AACAACT	ACGACTTGGAA	TGGGAGGCGCTCTGTCTGTGCGCATGCGACAGCAAGAAA	2007	
Qy	1684	GTCTTCTGT	CTCAAGAGTCACTGGG	CAAGTAATGACCAAAAAGATGTCCAGATTTCTACT	1743	
Db	2008	GTCACTGT	CTGCGAGTCAATGTGGG	CCCAAGTGGGACCCAAAAATGTCTGAAAGCCCT	2067	
Qy	1744	CGACTG	AAACTGTGCGCCCTGT	TTCTCCCATGMAAAAGACTGTATGTGACTGTCT	1803	
Db	2068	CGACTG	AAACTGTGAAGCCCTGT	GTCTCTTCTGTGAAGAGAGTGTATGTGACCCCA	2127	
Qy	1804	TTCACTGAGT	GGACACCCCTCCCAAGGAT	GTGCCAAGCAGAAATGCCACAGT---	1860	
Db	2128	TATAGT	GACTGGACATCATGGCCCT	CTTCTGTATGAAGAGGGACTCCAGTATCAGGAAG	2187	
Qy	1861	CAGTCTCGAT	PACAGAATCATCTCC	AGAGAGCCCAATGGAGCCAGGAATGCCCAGAT	1920	
Db	2188	CAGTCTAGGCAT	CGGGTCATCACTGCC	AGCCCAACGGGGCCGAGACTGCACAGAT	2247	
Qy	1921	ACCTTATAT	GAGGAGAGAGTGTG	RAAGATGTTTCCCTTGTGTCTGTATTCGGTGGAG	1980	
Db	2248	CCCCCT	TATGAAGAGAAGCCCTGT	GAGGCACCTCAGCGTGCCTCAAGAGTACAGGTGGAAG	2307	
Qy	1981	CCACAGAAAT	GGAGCCCTTGCATCTT	AGTGCCAGAGTCTCTGTGGCAGGGAATAACGGGC	2040	
Db	2308	ACTC	CAAAATGGCCAGATGCC	AATTAAGTCCCTTGGAGCGTGCACACAGACGCCCTTGA	2367	
Qy	2041	AGCAGT	GAAGCCTGTGGAAGGGG	TACAAACAGAGCTGTCTCATGCATCTCTGATGAC	2100	
Db	2368	GCACAG	GAAGGCTGTGGGCT	TGGCGCAGCGCAGAGGCCATTAATTTGTCCGCAAGCAAGAT	2427	
Qy	2101	AACCGGT	CAGCAGAAATGATG	GAATGCTCAAGCAGACAAAACGGCATGCTCTCCTTTGTG	2160	
Db	2428	GGAGGAC	AGGCTGGAAATCCATG	AGTGCTACAGTATGACGGCCCTGTGCGAGCCCTTACC	2487	
Qy	2161	CAAGAA	TGCACAGTCCCATGT	CGAGAAGACTGCACCTTTCACTGCTGGTCCAAAGTTACG	2220	
Db	2488	CAGCGCT	CGCAGATCCCTCG	CCAGGATGACTGTCAATTTGACCAAGCTGGTCCAAAGTTTCT	2547	

```
3289 ATTCCCTGCTTGGTGGATGCGTGGTCAACTGTCACTCTCAGGGTGGACGGCTTGGACA 3348
3628 ACCTCTGATGTTGGAATGCGTGGTCACTGTCAGCTTCTGATGGTCTCCTTGGTCA 3687
3349 GAGTGTTCACAGACCTGTGGCCATGAGGTGCGAATGAGCGGAGTCTGATTTATCATTTG 3408
3688 GAATGTTCTCAACATGTGGCTCACAGGAATAATGATCGGAAGACGACGTGACCCAG 3747
3409 CCAACCCAGGAGAGACGCGCCATGCCCAACAGAGCTTACCCAGGAGAAAACCTGCCCA 3468
3748 CCCTTTCAAGGTGATGGAAGACCATGCCCTTCCCTGATGGAACAGTCCAAACCCCTGCCCA 3807
3469 GTGACCCCTGCTACAGCTGGGTGCTTGGCACTGCTGTCATGTAATTTGGAGGTTGA 3528
3808 GTGAGGCTTGTATCGTGGCAATATGCCAGTGTCTCCATGCCAGTGCAGAGGCC 3867
3529 GACTGTGGGAAAGAGTTTCAGATCCGACGCTTTCCTGATGCTCCACAGTGTCCAAATA 3588
3868 CAGTGTGGAGAGGACACAGAAACAGGAACATTTCTGTGTAGTAAATGATGGGTGAGCT 3927
3589 TCTCATGCGAGCTGGAGCTGTCGAGGATGCACTGTGTGGAGAAATGCCCTT----- 3638
3928 GATGATTTGAGCAAAATGTTGGATGAGGAATTCGTGTGCAATTTGAATCAATATAGAT 3987
3639 -----TCAGGACAGCATCTGAGAGCAGCTGTGTTCTGTGCTTGGCCAGGAGACTGCCAT 3693
3988 GGTAATAAAAAATATGCTTCTGGAGGAATCCTGAGCCAGGCTTCCAGGTGACTGTTAT 4047
3694 TTAACAGATGTCAGATGAGGACACATGTGAATTAACCTGATGATGGAAGAGCTTTT 3753
4048 TTGAAGGACTGCTTCTTGGAGCCTGTGTGAGCTGACCTGTGTGAATGGTGAGGATCTA 4107
3754 GAGACTGTGGGCGCCAGCTAGATCAAGGACTTTTATAATTCAGTCTTTTGGAGAACCA 3813
4108 GGTGTTGGTGGATACAGTCACTGATCCAGCCGTTGATTTATACAGAACTAGAGATCAG 4167
3814 GACAGCTGCCCCCAACAGGTTCTAGAAAACGCCCCCTGTACAGGAGGCAAAATGTTATCAC 3873
4168 CATCTGTGCCAGAGCAGATGTTAGAAAACAAAATCATGTTATGATGGACAGTGTATGAA 4227
3874 TACACATGAAGAGCAAGCTTTGGACATACGAACTGATGATGTCGAGGTTCA 3933
4228 TATAATGATGCGCAGTCTTGGAGGCTCTTCCGAAACAGTGTGGTGTCAAAAGGTCA 4287
3934 GATGGCGTTAATGTACAGAGGCTGCTCCCTCAGGCCGCTCTGCTGCCATTCGGCAG 3993
4288 GATGGTATAAATGTAAACAGGGGCTGCTTGGTATGAGCCAGCTGATGCCAGAGTCT 4347
3994 TGCATTCAGCTGCGAGAAAACCTTCTCTACTGTACAGGGTGGAGTCTGTGGTTGT 4053
4348 TGTAAACCCACCTGTAGTCAACCCCACTCGTACTGTAGCGAGACAAAACATGCCATTGT 4407
4054 GAGAAGGGCTATACAGAGATAATGAATCAATGTTTCTCGATTACTTGCATGAAGTA 4113
4408 GAAGAAGGTACACTGAAGTCAATGCTTCTACAGACCCCTTGACCATGACACTTATC 4467
4114 CAGGCTCAGAGATATAAAGCTGATGTGAAAACCTTTCTGGAAAACAGAC 4168
4468 CCGTGGTGTATTACCCACCATGGAGGACAAAAGAGGAGATGTGAAAACCATGTC 4522
```

RESULT 13

ACCT72165

ID ACC72165 standard; DNA; 4881 BP.

CX ACC72165;

CX ACC72165;

CX ACC72165;

DT 07-JUL-2003 (first entry)

CX Human NOV47d coding sequence.

CX Human; NOV; antidiabetic; anorectic; antibacterial; virucide;

CX Human; NOV; antidiabetic; anorectic; antibacterial; virucide;

CX Human; NOV; antidiabetic; anorectic; antibacterial; virucide;

CX Human; NOV; antidiabetic; anorectic; antibacterial; virucide;

CX Human; NOV; antidiabetic; anorectic; antibacterial; virucide;

CX Human; NOV; antidiabetic; anorectic; antibacterial; virucide;

CX Human; NOV; antidiabetic; anorectic; antibacterial; virucide;

CX Human; NOV; antidiabetic; anorectic; antibacterial; virucide;

CX Human; NOV; antidiabetic; anorectic; antibacterial; virucide;

CX Human; NOV; antidiabetic; anorectic; antibacterial; virucide;

KW aniparkinsonian; antilipemic; gene therapy; metabolic disorder;

KW diabetes; obesity; infection; cachexia; cancer; dyslipidaemia;

KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;

KW immune disorder; haematopoietic disorder; gene; ds.

OS Homo sapiens.

XX WO2003029423-A2.

XX 10-APR-2003.

XX 02-OCT-2002; 2002WO-US031358.

XX 02-OCT-2001; 2001US-0326483P.

XX 05-OCT-2001; 2001US-0327342P.

XX 09-OCT-2001; 2001US-0327917P.

XX 09-OCT-2001; 2001US-0328029P.

XX 09-OCT-2001; 2001US-0328044P.

XX 09-OCT-2001; 2001US-0328056P.

XX 12-OCT-2001; 2001US-0328849P.

XX 15-OCT-2001; 2001US-0329414P.

XX 17-OCT-2001; 2001US-0330142P.

XX 22-OCT-2001; 2001US-0341058P.

XX 24-OCT-2001; 2001US-0339266P.

XX 29-OCT-2001; 2001US-0349575P.

XX 01-NOV-2001; 2001US-0346357P.

XX 12-APR-2002; 2002US-0371972P.

XX 12-APR-2002; 2002US-0371980P.

XX 17-APR-2002; 2002US-0373261P.

XX 19-APR-2002; 2002US-0373805P.

XX 23-APR-2002; 2002US-0374738P.

XX 16-MAY-2002; 2002US-0381101P.

XX 17-MAY-2002; 2002US-0381635P.

XX 29-MAY-2002; 2002US-0383830P.

XX 01-OCT-2002; 2002US-00262839.

XX (CURA-) CURAGEN CORP.

XX Alsobrook JP, Anderson DW, Boldog FL, Burgess CE, Catterton E;

XX Edinger SR, Ellerman K, Gerlach VL, Gorman L, Guo X, Ji W;

XX Kekuda R, Leach MD, Li L, Miller CE, Patturajan M, Rieger DK;

XX Rothenberg ME, Shimkets RA, Smithson G, Spytek KA, Taupier RJ;

XX Vernet CAM, Voss EZ, Zerhusen BD, Zhong M;

XX WPI: 2003-381525/36.

XX P-PSDB; ABR58453.

XX NOVX polypeptides and nucleic acids useful for diagnosing, preventing or

XX treating NOVX-associated disorders, e.g. diabetes, obesity, cancer or

XX dyslipidemia, and in chromosome mapping, tissue typing or

XX pharmacogenomics.

XX Claim 20; Page 256-257; 487pp; English.

XX The present invention relates to novel human NOV proteins and their

XX coding sequences (ACC72075-ACC72181 and ABR58363-ABR58469). The NOV

XX proteins are useful in manufacturing a medicament for treating a syndrome

XX associated with a human disease. The NOV proteins and coding sequences

XX may be used to diagnose, treat or prevent metabolic disorders such as

XX diabetes or obesity, infections, cachexia, cancer, neurodegenerative

XX disorders such as Alzheimer's disease or Parkinson's disease, immune

XX disorders, haematopoietic disorders and various dyslipidaemias

XX Sequence 4881 BP; 1391 A; 1116 C; 1320 G; 1054 T; 0 U; 0 Other;

XX Query Match 26.8%; Score 1177.4; DB 7; Length 4881;

XX Best Local Similarity 57.3%; Pred. No. 0;

XX Matches 2439; Conservative 0; Mismatches 1711; Indels 105; Gaps 13;

XX 10 TGCAATCAGAGCTGAACCGAAGTGGTTGCAATGAATATGCAACACTTGGCCCTT 69

XX 289 TGCATCCAGAAACAAAGACATCTCTCGGAGGATATCATCTGTGAGTACTTTGAGCCC 348

QY 70 CAGCCTCTACAGAACAGGCTTGCTCAATTCCTTGTCCCGGAAATGTGTAGTATCTGAG 129  
Db |||||  
QY 349 AAGCCTCTCTGGAGCAGGCTTGCTCAATTCCTTGCAGCAAGATTGCATCGTGTCTGAA 408  
Db |||||  
QY 130 TTCTTACCATGGTCCAACTCTAGCAAGGGATGTGGGAAGAAATTCAGCAGATAGAACTCGC 189  
Db |||||  
QY 409 TTTTCTGCTGGTCCGAATGCTCCAAGACCTGGCGCAGCGGGCTCCAGCACCGCGCGT 468  
Db |||||  
QY 190 GCGGTCAATAGTCCCTCTCTCTTGTGGTGTTCGATGTCCAAATCTGACTGAGTCAAGA 249  
Db |||||  
QY 469 CATGTGGTGGCCCGCCGAGTCTCGAGGCTCTGGCTGTCTCAAACCTTCAGCGAGTTCCAG 528  
Db |||||  
QY 250 GCCTGTGATGCTCCCAATTCCTGTCTCTTGGGAAGAGGAATATACATTTAGCCTTAAG 309  
Db |||||  
QY 529 GTGTG-----CCAATCAAGTCCATGCGAGGCCGAGGCTCAGGTACAGCCTGCAT 579  
Db |||||  
QY 310 GTTGGACATGGAGTAATGAGACTGTCTCATCTTAAAGAAATTAATCCAAAGGGAAGA 369  
Db |||||  
QY 580 GTGGGGCCCTGGAGCACCTGCTCAATGCCCCACTCCCGACAAGTAAAGCAAGCAAGAGA 639  
Db |||||  
QY 370 ACTGTTCTGGATTTTAA-----CTCTGATTCAAATGAG 402  
Db |||||  
QY 640 CGCGGAAGATAAAGAACGGGAAGAACCGCAGCAAGAGGTAAAGGATCCAGAGCC 699  
Db |||||  
QY 403 CGAGTCACCTTTAAACATCAAAAGTTACA-----AAGCACATCATCATTCGAAGTCT 453  
Db |||||  
QY 700 CGCGAGCTTATTAAGMAAAGAGAAACAGAAACAGGCAAGACAGACAAGAGAACAAATAT 759  
Db |||||  
QY 454 TGGCAATAGAGATAGGTATCAACACCGCGAGCTTCGTGTCAAGAAATGATGGACAA 513  
Db |||||  
QY 760 TGGGACATCCAGATTGGATAACAGACAGAGAGGTATGTGCATTAACAAGACGGGAAA 819  
Db |||||  
QY 514 AATGCTATGTTAAGCCTTTGCTTCAAGATTCCTTCCATTTGACTGTTCAGTCTCTGCATC 573  
Db |||||  
QY 820 GCTGTGATTTAAGCTTTTCCAGCAAGAGAAGCTTCCATGACCTTCCAGTCTCTGTGTG 879  
Db |||||  
QY 574 ATGCCAAAGACTGTGAACCTCCAGTGTCTCTGGAGCCCTCTGCTCAAGACATGC 633  
Db |||||  
QY 880 ATCAACAAAGAGTCCAGGTTTCCAGTGTGTGAGAGTGGAGCCCTCTCTCAAAAACATGC 939  
Db |||||  
QY 634 CGTTCAGGAGTCTCTCCAGGATTTAGGAGCAGGAGCCCGGAACGTGAAGACATCGCT 693  
Db |||||  
QY 940 CATGACATGGTGTCCCTGCGAGGCACTGCTGTAGGACACGAACCATCAGCAGTTTCCC 999  
Db |||||  
QY 694 ATTGAGGTGGAAGAGATGTCTGAACTTCTTGAGAAAGAGGCTCGATTTGTTGAAGGA 753  
Db |||||  
QY 1000 ATTGGCAGTGAAGAGAGTGTCCAGATTTGAAGAAAAGAACCCCTGTTGTCTCAAGGA 1059  
Db |||||  
QY 754 GAACTTCTGCAGCAATGCTCCAGGTATTCCTGGAGAACTTCTGAATGGAAGAAATGCCAA 813  
Db |||||  
QY 1060 GATGGAGTTGCTCCCTGTGCCACGTAATGGCTGGAGAACTACAGAGTGGACTGAGTGCCTG 1119  
Db |||||  
QY 814 GTCTCTCTCTCTCGAGCAGCAGGATCCCACCTGGCATGTGACGGGACCGCTGTGTGGC 873  
Db |||||  
QY 1120 GTGGACCCTTTGCTCAGTCAGCAGCAAGAGGCGCGCAACCAAGACGGCCCTCTGTGGA 1179  
Db |||||  
QY 874 GGTGGATCCAGACCGCGAGGTGTACTGTGCCAG-----AGGTACCAAGCA 921  
Db |||||  
QY 1180 GGGGCGATCCAGACCGGAGAGTGTACTGCGTGCAGGCGCAACGAAAACCTCTCTCAAA 1239  
Db |||||  
QY 922 GCTGCGCACTGAGGCGCAAGAGTCTCTAGACCTGTGGAAAAGGCAATTTATGTGTGGGA 981  
Db |||||  
QY 1240 TTAAGTACCCACAAGAACAAAGAGGCTCAAGGCCAATGGACTTTAAATATATGACCTGGA 1299  
Db |||||  
QY 982 CCGGCGCGCTTCCCTCTCAGCTCTGGAATATCCCTTGTCTCTACGACTGATAGTATCT 1041  
Db |||||  
QY 1300 CCTATCCCTTAATACACAGCTGTGGCACAATTCCTTGTCCAACGTGAATGTGAAGTTCA 1359  
Db |||||  
QY 1042 TCCTGCTCAGCCTGGGCTGTGCATCCATGAAACTGTCAATGAACTTCAGACCTCAGGCGAAAAA 1401  
Db |||||  
QY 1360 CCTTGGTCAAGCTTGGGACCTTGACTTATGAAAATCTGATGATCAGCAAGGGAAGAAA 1419  
Db |||||

QY 1102 GGATTTAGAACGAGCAGCGCC-----ATGTCTCATGGAATCTACAGGCGCTGCA 1152  
Db |||||  
QY 1420 GGTCTCAAATCTGAGAGCGCGCATTAACCAATGAGCCACTGAGGCTCTGGGTAACC 1479  
Db |||||  
QY 1153 GGGCAATGCCCTCAATTTGGTGGATCTGTCTTCTTGTGAGGATCCAATGTGCTACCGATGG 1212  
Db |||||  
QY 1480 GGAACCTGCCCTCACTTACTGGAAGCAATTCCTGTGAAGAGCCTGCTGTATGACTGG 1539  
Db |||||  
QY 1213 ---CTGGCATCAGAAGGATCTGTTTCCCTCATATGGA---AATGTGGCCTGGGACAT 1266  
Db |||||  
QY 1540 AAGCGGTGAGACTGGGAGACTGCGAGCCAGTAAACGGAAGGAGTGTGTTCCAGGCAAG 1599  
Db |||||  
QY 1267 GGTATTTCTGAAGCGCTCTCCAGAAATGACCGCGAAGATGTATCAGGAGTCTTTGC 1326  
Db |||||  
QY 1600 CAAATTCAAGAGTGTGTGCATCAACAGTGTATGAGAAGAGTGTGACAGCAGCTGTGC 1659  
Db |||||  
QY 1327 CCAGTTCGCCCTCTCCCTGAGGAGGATCTTGTGAATTCCTCCGAAATGCACTGTGTG 1386  
Db |||||  
QY 1660 AGAGATGCCATCTTCCCATCCCTGTGSCCTGTATGCCCCCATGCCCGAAGACTGTGTG 1719  
Db |||||  
QY 1387 CTGAGCGAGTGGACGAGTGTGTCATCTCTGTTCCAGTCTCTTCAAAATAAAATCTCAGAT 1446  
Db |||||  
QY 1720 CTCAGCACATGCTCTACGTGTCTCTCTCTCTCACACCTCTCAGGGAAAACGACAGAA 1779  
Db |||||  
QY 1447 GGGAAACAGACCAAGTCAAGAACTATCTCTGSC---ACTGGCTGGGAGGTGGAAAGCCA 1503  
Db |||||  
QY 1780 GGGAAACAGATACGAGCAGATCCATCTGSCCTATGCGGGTGAAGAGGTGGAAATTCGC 1839  
Db |||||  
QY 1504 TGTCTCCCTTAGTCAGGCTCTCCAAGAGCATGTTTGTGTATATGACCATTCCTGTATGAG 1563  
Db |||||  
QY 1840 TGTCCAATAGCAGTGTCTTGCAGAACTACGAAGCTGTATGAGCATCTCTTGACAGTG 1899  
Db |||||  
QY 1564 CTTCTCTGGAGACATCGCCTTGGGCGCTTGTCTTGTAGGACACATTTGTTAACTGCGCTT 1623  
Db |||||  
QY 1900 TACCCTGCAAACTGTGTCCTGGGCGCAGTGCAATGAGGACACCTCAGTATCGTCTTTC 1959  
Db |||||  
QY 1624 AATGCAACCATTTGGTGGAAATGGAAGCCAGCTGTGTGTAGGCAATTCAGACTCGGAGA 1683  
Db |||||  
QY 1960 AACAACAATCAAGTGTGAATGGGAGGCTCTCTGCTGTGCGCATGCGACAGAAATAA 2019  
Db |||||  
QY 1684 GTCTTCTGTCTCAAGATCAAGTGGGACAAAGTAATGACAAAGATGTCAGATTTCTACT 1743  
Db |||||  
QY 2020 GTCACTGTGTGCGAGTCAATGTGGGCAAGTGGGACCCAAAAAATGTCCTGAAAGCCTT 2079  
Db |||||  
QY 1744 CGACCTGAACTGTGCGCCCTGTTTCTCCATGSCAAAAGAGCTGATTTGTGACTGCT 1803  
Db |||||  
QY 2080 CGACCTGAACTGTGAGGCTTGTCTGCTTCTTGTAAAGAGACTGTATTTGTGACCCCA 2139  
Db |||||  
QY 1804 TTCAGTGTGACACCCCTGCGCAAGGATGTGCCAAGCAGGAAATGCCACAGT---AAAA 1860  
Db |||||  
QY 2140 TATAGTGTGACATCATGCCCCCTCTTCTGTTAAAGAGGGACTCCAGTATCAGGAAG 2199  
Db |||||  
QY 1861 CAGTCTCGATACAGATCATCATCAAGAGCCATGAGGAGCCAGGATGCCAGAT 1920  
Db |||||  
QY 2200 CAGTCTAGCAGTCTGAGTCTGAGGCTTGTCTGAGGCACTCAAGCGTGCAGGCTACAGTGG 2259  
Db |||||  
QY 1921 ACCTTATATGAGCAGAGAGTGTCAAGATGTTTCTTGTCTCTGTATATCGGTGGAAG 1980  
Db |||||  
QY 2260 CCGCTCTATGAGAGAGGCTGTGAGGCACTCAAGCGTGCAGGCTACAGTGGGAAG 2319  
Db |||||  
QY 1981 CCAAGAAATGAGCCTTGTGATCTTGTGCGAGAGTGTGCTGCGAGGAAATACCGGC 2040  
Db |||||  
QY 2320 ACTCAAAATGGCGAGATGCAATTAGTCTTGGAGCGTGCAACAGACAGCCCTGGA 2379  
Db |||||  
QY 2041 AGCAGTGAAGCCTGTGGAAGGGGTTACAAACAGAGCTGTCTCATGCTCTCTCATGAC 2100  
Db |||||  
QY 2380 GCACAGAAAGCTGTGGGCTGTGGCGCAGGCAAGGCAATTAATTTGCGCAAGCAGAT 2439  
Db |||||  
QY 2101 AACCGCTCAGCAAGAAATGATGGAATGCTCTCAAGCAGACAAACGGCATGCTCTCTTGTG 2160  
Db |||||  
QY 2440 GGAGGACAGCTGGAATCCATGAGTGTCTTACAGTATGCAAGGCCCTGTGCGAGCCCTTACC 2499  
Db |||||  
QY 2161 CAAGATGACAGTCCCATGTCGAGAGACTGCACCTTCACTGCTTGGTCCCAAGTTTACG 2220  
Db |||||

2500 CAGGCGTCAGATCCCTGCGCAGGATGACTGCAATTGACAGCTGTCACAAAGTTTCT 2559  
2221 CCTGCTCCACGACGTGTAAGCCCAAAAGATAGGCGGCGACAGCTCACAGGGAAGC 2280  
2560 TCATGCAATGGAGACTGTGTGAGTGTAGGACAGAAAGCGACCTCTTGTGGAAAAAGT 2619  
2281 AGAAGAGGAGAAATGCCAGGATCTGACCTTTACCTCTAGTGGAGACAGAACTATGT 2340  
2620 AAAAGAGGAAAAATGTAATAATTCCTATTCCTCTGATTCAGTTCAGTATGT 2679  
2341 CCTGTGATGAATTTATATCCCAACCTTATGAAACTGTGTAGATGATTCCTCCAGAA 2400  
2680 CCTGTGACAAATATATGCAACACCTGTGGGAACTGTGTACAGCTGATTTTACAGAG 2739  
2401 GGCAGAAAGGAGCCTCACCGAGACTGCGGTACAGACAGACAGCAAAAGATGTGGAGAA 2460  
2740 GGAAGAGTGAAGTGTGCTGGGATGAAAGTACAAGGAGACATCAAGGAATGCGGACAA 2799  
2461 GGCCTGCGCTTTGAGCAGTAGCTGTTCTGTATAAAATGGAAGACCTGTTGACCCCTCC 2520  
2800 GGATATCGTTTACCAAGCAATGGCATGCTACGATCAAAATGGCAGGCTTGTGGAAACATCT 2859  
2521 TTCTGACAGAGCTCTGCTTACATTCAGAAATAATGTCTATTCCTGCTGCCATTTGATTGC 2580  
2860 AGATGTAACAGCCATGTTTACATTTAGGAGGCTGCATCATCCCTGCCCTCAGACTGC 2919  
2581 AAGTTAAGCGATGCTAGTTAGTTGGGGTCTTCAGTTTCACTTTGTGGAATGAGTGAGA 2640  
2920 AAGCTCAG\*GAGTGTGCTCAACTGCTCGCTGCGCTGCGCAAGTCTGTGGAGTGGTGAAG 2979  
2641 ATTCGATCCAAATGGCTTAAAGAAACCTTCAATGGAGGACGACCATGTCCCAACTG 2700  
2980 GTTCGTCTAAATGGCTGCGTGAAAAACCATATAATGGAGGAAGGCTTGTGCCAAACTG 3039  
2701 GATC-----TCAAGAAATCAGGTACATGAGGAGTCCCATGTTACAGTGAAGTCAATCAG 2754  
3040 GACCATGTCAACAGGACAGAGTGTATGAGTGTGCTCCATGCCACAGTGAATGCAACCG 3099  
2755 TATTCCTGGTTTGAACAACCTGCTTCAATGCAAAATCAAAATGAGTGAAGTCCCTG 2814  
3100 TACCTATGGGTACAGAGCCCTGAGCATCTGCAAGGTGACCTTTGTGAATATGCGGGAG 3159  
2815 CGCTGTGGAGGAGAAACAATCTAGGAAATCAGATGTGT---GAATCTCGGATGT 2871  
3160 AACTGTGGAGGCGCTGAAACCCGAAAGTGAATGATGATGCAAGATACACAGATGCG 3219  
2872 GAAGGTGGAGCAGTGAATAGCAACCTGTGCAACAGGATGAAATTCGCCAGAAACCCAG 2931  
3220 CCTTCTGAAACATGATAGGATTAACCTCTGTGACCCAGAAAGATGCCCCCTGGCTCTAGA 3279  
2932 TCCTGTTCTTTATGTTCCCATGAGTGTGTCATGCTGAGTGGGACTTTTGGAGCAAA 2991  
3280 GTGTGCAAAATTCATGCTCCCTGAGGACTGTGTGATATCTGAAATGGGGTCCATGGACCCAA 3339  
2992 TGCCCCCAGTCACTGCGATCCCAACAAT---GCAGAGAAAGTCCGCCACCTGTCTAAGA 3048  
3340 TGTGTTTTCCTTTCATCAAGCAGTTTCCGCGCAAGGTGAGTATCCCATCAGACAA 3399  
3049 CCATCACTGACATCAGGACTGTGTGTAAGACTCACAGGTGACGCTTGCCTCTGTAAT 3108  
3400 CCAGCTGATGAAGGAAGATCTTCCCTTAATGCTGTTGGAAGAAAGAACCTGTAACTGAA 3459  
3109 GAAATATGCTTCCAGTTCAGATCAATCTAACAGAGTGGAGACATGCCAGCTGAGTGAA 3168  
3460 AAAAACTGCTACCATGATGATATATGTAACAGACTGGAGTACATGTCAGTGTGAGTGAG 3519  
3169 AACGCACTGTGGTCAAGGCGTCAGGAACCCGCTCTGCTAAGCTGTGTGTGTCAGTGTGCG 3228  
3520 AAGGCACTTTGTGGAATGGAATAAAAAACAAGGATGTGTGAATGTTGTGTTGAAAGTATGCG 3579  
3229 AAGCCAGTCAGCATGGAACCAATGTGAGCAGCAATAATTTGAGAGAGGCCCCAGAGATGAGC 3288

Db 3580 AAGTCAGTTGACCTGAAATATTTGTGAAGCGCTTGGCTTGGAGAAAGTAATGGCAGATGAAC 3639  
QY 3289 ATTCCCTGCTTGTGGAATGCGTGTCAACTGTCACTGTCACTGTGAGGTGGAGCGCTTGCACA 3348  
Db 3640 AGTCTCTGCAATGTGGAATGCCCTGTGAATGTCACTGTCACTTCTGTATGGTCTCCTTGGTCA 3699  
QY 3349 GAGTGTTCACAGACCTGTGGCCATGGAGGTGCAATGAGCGGACTCGATTTATCATTTATG 3408  
Db 3700 GAATGTTCTCAACATGTGGCTTCAAGGAAAAATGATCCGAAGACGAACAGTGAAGCCAG 3759  
QY 3409 CCACCCAGGAGAGGAGCGGCATGCCCCACAGAGCTTACCAGGAGAAACCTGCCCA 3468  
Db 3760 CCCTTTCAAGGTGATGAAGACCATGCCCTTCCCTGATGACAGCTCCAAACCTGCCCA 3819  
QY 3469 GTGACCCCTCTGACAGCTGGGTCTTGGCAATGTGTCTGTGATGTAATTTGGAGGTGGA 3528  
Db 3820 GTGAAGCCTTGTATCGGTGCAATATGGCAGTGTCTCCATGCAAGTGCAGGAGGCC 3879  
QY 3529 GACTGTGGGAAGAGATTAGATCCGAGCCTTCTGTGATGCTCCACAGTGTCCAGTGTCAATA 3588  
Db 3880 CAGTGTGGAGAGGAGCCAGAACCAAGAACATTTCTTGTGTAGTAAGTATGATGGGTGAGCT 3939  
QY 3589 TCTCATGCACTGGAACGTGTGCGAGGATGCACTGTGTGGAGAAATGCCCTT----- 3638  
Db 3940 GATGATTTCAAGAAAGTGTGTGATGAGAAATCTGTGCTGACATTTGAATCATTATAGAT 3999  
QY 3639 -----TCAGGACAGCATCTGGAAGCAGCTGTGTCTGTGCTTGGCCAGGAGACTGCCAT 3693  
Db 4000 GGTAAATAAAATATGTTCTGTGAGGAATCTCGAGCAGCGCTTGGCCAGGTGATTTAT 4059  
QY 3694 TTAACAGAAATGTGTGAGTGGAGCAGATGTAATTAACCTGCAATTTGATGGAAGACTTT 3753  
Db 4060 TTGAAGGACTGTGCTTCTGAGCCTGTGTGACGTGACCTGTGTGATGATGAGTGTGAGTACTA 4119  
QY 3754 GAGACTGTGGCCCGCAGTCTAGATCAAGCACTTTTATAATTCAGTCTTTTGTGAACCAA 3813  
Db 4120 GGCCTTGTGGAATACAGTCCAGTCCAGACCGGTGATTATACAAGAACTAGAGAAATCAG 4179  
QY 3814 GACAGTGTCCCCACAGGTTCTAGAAACAGCGCTTGTACAGAGGCAATTTATTCAC 3873  
Db 4180 CATCTGTGCCAGAGCAGATGTTAGAAACAAATCATGTTATGATGGACAGTGTCTATGAA 4239  
QY 3874 TACATGGAAGCAAGTCTTTGGAAACAATAAGCAAGAACTGTATGGTCCAGCGTTCA 3933  
Db 4240 TATAAATGGATGGCCAGTGTCTTGGAGGGCTCTTCCGACAGTGTGTCTCAAAGTCA 4299  
QY 3934 GATGCGCTTAAATGTCAAGAGGCTGCTCCCTCAGGCCGTCTGTGCTGCAATTCGSCAG 3993  
Db 4300 GATGTTATAAATGTAAACAGGGGCTGCTTGTGATGAGCCAGCTGATGCCAGGCTCT 4359  
QY 3994 TGCATTCCAGCTCGAGAAACCTTTCTCTCTGTACTGTACAGGGTGGAGTCTGTGGTTGT 4053  
Db 4360 TGTAACCCACCGTGTAGTCAACCCCATCTGTACTGTAGCGAGCAAAAACATGCCATTGT 4419  
QY 4054 GAGAAGGCTTATCAGAGATAAAGAAATCAAAATGAGTGTCTGTGATCTGATGTAAGTA 4113  
Db 4420 GAAGAGGTTACACTGAAGTCAATGCTTCTTAACAGCAGCCCTTGGACAAATGCACACTATC 4479  
QY 4114 CCAGGCTCAGAGGATAAAGAGCTGATGTAATAAATCCTTTCTGGAATAACAGAC 4168  
Db 4480 CCGTGTGTGTTATTACCACCATGGAGACAAAGAGGAGATGTGAAAACAGTTC 4534

## RESULT 14

ACC72166

ID ACC72166 standard; DNA; 4679 BP.

XX ACC72166;

XX 07-JUL-2003 (first entry)

XX Human NOV47e coding sequence.

XX

KW Human; NOV; antidiabetic; anorectic; antibacterial; virucide;  
 KW immunomodulator; cyrostatic; nootropic; neuroprotective;  
 KW anticancer; antileukemic; gene therapy; metabolic disorder;  
 KW diabetes; obesity; infection; cachexia; cancer; dyslipidaemia;  
 KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
 KW immune disorder; haematopoietic disorder; Gens; ds.

XX Homo sapiens.

XX WO2003029423-A2.

XX 10-APR-2003.

XX 02-OCT-2002; 2002WO-US031358.

XX 02-OCT-2001; 2001US-0326483P.

XX 05-OCT-2001; 2001US-0327342P.

XX 09-OCT-2001; 2001US-0327917P.

XX 09-OCT-2001; 2001US-0328029P.

XX 09-OCT-2001; 2001US-0328044P.

XX 09-OCT-2001; 2001US-0328056P.

XX 12-OCT-2001; 2001US-0328849P.

XX 15-OCT-2001; 2001US-0329414P.

XX 17-OCT-2001; 2001US-0330142P.

XX 22-OCT-2001; 2001US-0341058P.

XX 24-OCT-2001; 2001US-0339286P.

XX 24-OCT-2001; 2001US-0343629P.

XX 29-OCT-2001; 2001US-0349575P.

XX 01-NOV-2001; 2001US-0346357P.

XX 12-APR-2002; 2002US-0371972P.

XX 12-APR-2002; 2002US-0371980P.

XX 17-APR-2002; 2002US-0373261P.

XX 19-APR-2002; 2002US-0373805P.

XX 23-APR-2002; 2002US-0374738P.

XX 16-MAY-2002; 2002US-0381101P.

XX 17-MAY-2002; 2002US-0381635P.

XX 29-MAY-2002; 2002US-0383830P.

XX 01-JUN-2002; 2002US-00262839.

XX (CURA-) CURAGEN CORP.

XX WPI; 2003-381625/36.

XX P-PSDB; ABR58454.

XX NOVX polypeptides and nucleic acids useful for diagnosing, preventing or  
 XX treating NOVX-associated disorders, e.g. diabetes, obesity, cancer or  
 XX dyslipidemia, and in chromosome mapping, tissue typing or  
 XX pharmacogenomics.

XX Claim 20; Page 257-258; 487pp; English.

XX The present invention relates to novel human NOV proteins and their  
 XX coding sequences (ACC72075-ACC72181 and ABR58363-ABR58469). The NOV  
 XX proteins are useful in manufacturing a medicament for treating a syndrome  
 XX associated with a human disease. The NOV proteins and coding sequences  
 XX may be used to diagnose, treat or prevent metabolic disorders such as  
 XX diabetes or obesity, infections, cachexia, cancer, neurodegenerative  
 XX disorders such as Alzheimer's disease or Parkinson's disease, immune  
 XX disorders, haematopoietic disorders and various dyslipidaemias

XX Sequence 4679 BP; 1331 A; 1051 C; 1268 G; 1029 T; 0 U; 0 Other;

XX Query Match 25.8%; Score 1133.4; DB 7; Length 4679;

XX Best Local Similarity 57.7%; Pred. No. 0;

XX Matches 2332; Conservative 0; Mismatches 1606; Indels 105; Gaps 13;

Db 288 TGATCCAGAAAGACAAAGACATTCCTGGGAGGATATCATCTGTGAGTACTTTGAGCCC 347  
 QY 70 CAGCCTCTCAGAAACAGGCTTGCTATCTCTTGTCCCGGATTGTGTAGTATCTGAG 129  
 Db 348 AAGCCTCTCTCCGGAGCAGGCTTGCTCTCATCTCTTCCAGCAAGATTGATCGTCTGAA 407  
 QY 130 TTCTTACCATTGTCCTCAACTGTAGCAAGGATGTGGAGAAATTCAGCATAGAACTCGC 189  
 Db 408 TTTTCTGCTGCTCGAATGCTCCAGAGACCTGCGGAGCGGCTCCAGCACCGAGCGGT 467  
 QY 190 GCGGTCAATAGCTCCCTCTCTTTGGTGGTTGCAATGTCCAAATCTGACTGAGTCAAGA 249  
 Db 468 CATGTGGTGGCGCGCCCGCAGTTCGGAGGCTCTGGCTGTCCAAACCTGACGGAGTTCAG 527  
 QY 250 GCTGTGATGCTCCCATTTCTCTCTTGGGAGAGAGATATACATTTAGCCTTTAAG 309  
 Db 528 GTGTG-----CCAATCCAGTCCATGCGAGCGCGAGGCTCAGTACAGCTGCAT 578  
 QY 310 GTTGACCAATGGAGTAAATGAGACTGCTCATTTAAAGAAATTAATCCAAAGCGGAGA 369  
 Db 579 GTGGGGCCCTGGAGCACCTGCTCAATGCCCTCCGACCAAGTAAGACAAGCAGGAGA 638  
 QY 370 ACTGTTCTGGATTTTAA-----CTCTGATTCAATGAG 402  
 Db 639 CCGGGAAGATTAAGAACGGGAAAGGACCGCAGCAAGGAGTAAGGATCCAGAGCC 698  
 QY 403 CAGGTACCTTTTAAACATCAAAAGTTTACA-----AAGCACATCATCTCGAAGTCT 453  
 Db 699 CGCGAGCTTTATTAGAAAAAGAGAAACAGAAACAGGACAGACAGACAGAAACAATAT 758  
 QY 454 TGGGCAATAGATAGGTATCAACCCGCGAGGTTTCTGTACAAAGAGTATGATGACAA 513  
 Db 759 TGGGACATCCAGATTGGATATCAGACACAGAGGTTATGTGATTAACAGACGGGAAA 818  
 QY 514 AATGCTATGTTAAGCCTTTGCTTCAAGATTCCTTCCCATGACTGTTTCAAGTCTGATC 573  
 Db 819 GCTGCTGATTAAAGCTTTTGGCAGCAAGAGAGCTTCCAATGACCTTCCAGTCTGTGTG 878  
 QY 574 ATGCCCAAGACTGTGAACCTCCAGTGGTCTCTCTGGAGCCCTGCTCCAGACATGC 633  
 Db 879 ATACCAAAAGATGCGAGGTTTCCAGAGTGTTCAGAGTGGAGCCCTGCTCAAAAACATGC 938  
 QY 634 CGTTCAGGGAGTCTCTTGGCCAGGATTTAGGACAGGAGCGGAAACGTTGAAGCAGTGGT 693  
 Db 939 CATGACATGCTGCTCCCTCGAGGACTCGTGTGAAGACAGCAACCATCAGGCAGTTCC 998  
 QY 694 ATTGAGGTGAAAGAGAGTCTCTGAACTTCTTGAAGAAAGGCTGCTGATTTGAAGA 753  
 Db 999 ATTGGCAGTGAAGAGAGTGTCCAGAAATTTGAAGAAAAAGAACCTGTTGTCTCAAGA 1058  
 QY 754 GAACTTCTGAGCAATGTCCTCCAGGATTTCTGGAGAACTTCTGAATGGAAGAAATGCCAA 813  
 Db 1059 GATGGAGTGTCCCTGTGCGCAGTATGGCTGGAGACTTACAGAGTGGACTGAGTGGCT 1118  
 QY 814 GTCTCTCTCTCTGAGCAGGAGTATCCCACTGGCATGTGACGGGACCCGCTGTGTGGC 873  
 Db 1119 GTGGACCTTTGCTCAGTCAGCAGGACAAAGAGCGCGCAACAGACAGCGCCCTCTGTGGA 1178  
 QY 874 GTGGGATCCAGACCGGAGGAGTGTACTGTGCCCAG-----AGCGTACAGCA 921  
 Db 1179 GGGGGCATCCAGACCGGAGAGGTTACTCGTGGAGGCAACGAAAACTCTCTCACAA 1238  
 QY 922 GCTGCGCAGTCTGAGGCGCAAGGAAGTCTCTAGACTGTGGAAAAAGGCAATATGTGGGA 981  
 Db 1239 TTAAGTACCACCAAGAAACAAAGAGGCTCAAAGCCAAATGGAATTAATAATGCACTGGA 1298  
 QY 982 CCGCGCGGTTGCCCTCTCAGCTTGCATATCCCTTGTCTTACGAGACTGCATATCTCT 1041  
 Db 1299 CCTATCCCTAATACTACAGCTGTGCCCATTCCTTGTCCCACTGAATGTGAAGTTTCA 1358  
 QY 1042 TCCTGTGTCAGCTGGGGCTCTGTCATCCATGAAACTGTCAAGAACCTCAGGGGAAAAA 1101



b 1359 CCTTGTGACGCTTGGGACCTTGTACTTATGAAAACACTGATGATCATGCAAGGGAAGAAA 1418  
y 1102 GGATTTAGAACGAGGACGG-----CCATGTCCTCATGGAATCTACAGGCGCTGCA 1152  
b 1419 GGCCTTCAAACTGAGGAAGCGCGCAATTACCAATGAGCCCACTGGAGGCTCTGGGGTAACC 1478  
y 1153 GGGCAATGCGCTCATTTGGTGAGTCTGTTCCTCTGTGAGGATCCAAATGTGTACCGATGG 1212  
b 1479 GGAACCTGCGCTCATTTACTGAGGCCATTCCTGTGAAGCCCTGCTGTATGACTGG 1538  
y 1213 -----CTGGCATCAAGAGGATCTGTTTCCCTGATCATGGA-----AATGCGCTGGGACAT 1266  
b 1539 AAAGCGGTGAGACTGGGACCTGCGAGCCAGATAACGGAAGGAGCTGTGGTCCAGGACG 1598  
y 1267 CGTATTTCTGAAGCGCTCTGCGAGAAATGACCGGAGAAAGATGATATCAGGAGTCTTTGC 1326  
b 1599 CAAGTTCAAGAGGTTGTGTGCACTACAGTGTATGAGAGAGAGTTGACAGACAGCTGTGC 1658  
y 1327 CCAGTTCCCGCTCTCTCTGAGAGGAAGTCTTGTGAAATTCCTGCGGAATGAGACTGTGTG 1386  
b 1659 AGAGATGCCATCTTCCCATCCTGTGGCTGTGATGCCCATGCCCAAGACTGTGTG 1718  
y 1387 CTCAGCGGTGACGAGTGTCTATCCCTGTCCTGTCAGTCTGTTCAATAAATAAATCAGAT 1446  
b 1719 CTCAGCACATGGTCTACGTGGTCTCTCTCTCACACACCTGCTCAGGGAAGACAGACAA 1778  
y 1447 GGAACACAGACACGAGTCAAGAACTATCTTGGC---ACTGGCTGGGGAAGGTGGAAGCCA 1503  
b 1779 GGAACACAGATACGACACGATCCATCTTGGCTATGCGGGTGAAGAGGTGGAATTCG 1838  
y 1504 TGTCCCGCTAGTCAGGCTCTCAAGAGCATGTTGTGTGAATGACCATTCCTGTATGAG 1563  
b 1839 TGTCCAATAGCAGTCTTTGCAAGAAGTACGAAGCTGTAATGAGCATCTTGTGCAAGTG 1898  
y 1564 CTTCACTGGGAGACATCGCTTGGGCGCTTGTCTGAGGACACATTTGTAACTCCCTT 1623  
b 1899 TACCCTGGCAAACTGGTCCCTGGGCGAGTGCATTTGAGGACACCTCAGTATCGTCTTC 1958  
y 1624 AATGCAACCATTTGGTGAATGAGAGAGCAAGTGTGTGTGATGAGCATTCAGACTCGGAGA 1693  
b 1959 AACACAACTACGACTTTGGAATGGGAGGCTCTCTGCTCTGTGCGGATGCAAGAA 2018  
y 1684 GTCTTCTGTCAAGAGTCAGTGGGACAAAGTAATGACCAAAAGATGTCCAGATCTACT 1743  
b 2019 GTCATCTGTGTGCGATCAATGTGGGCGAAGTGGGACCCAAATAATGTCTGAAGCCTT 2078  
y 1744 CGACCTGAAACTGTGCGCCCTGTTTCTCCCATGCAAAAAAGACTGTATGTGACTGCT 1803  
b 2079 CGACCTGAAACTGTAAGGCTTGTCTGCTTCTTGTGAAGAGACTGTATTTGTGACCCCA 2138  
y 1804 TTCAAGTGGTGGACACCTGCGCAAGGATGTGCAAGAGGAAATGCCACAGT---AAAA 1860  
b 2139 TATAGTGACTGGACATCATGCCCCCTTTGTGTGAAGAGGGGACTCCAGTATCAGGAG 2198  
y 1861 CAGTCTCGATACAGAAATCATCATCCAAAGACAGCCAAATGGAGGCGAGAAATGCCAGAT 1920  
b 2199 CAGTCTAGGCACTCGGTTCATCATTTAGCTGSCCAGCCAAAGGGGCGGAGACTGCAAGAT 2258  
y 1921 ACCTTATATGAGAGAGAGTGTGAAGATGTTTCTTGTGCTCTGTATATCGTGGAG 1980  
b 2259 CCCTCTATGAAGAGAGGCGCTGTGAGGCACTTCAAGCGTGGCCAAAGCTACAGGTGGAAG 2318  
y 1981 CCACAGAAATGGAGCCCTTGCATCTTAGTGCCAGAGTCTGTGGCAGGGAATAACGGGC 2040  
b 2319 ACTCACAATGGCGCAGATGCCAATTAGTCTTGGAGCGTGCACAAAGACAGCCCTGGA 2378  
y 2041 AGCAGTGAAGCTGTGGAAGAGGGTTCAAAAGAGAGTGTCTCATGCAATCTCTGATGAC 2100  
b 2379 GCACAGGAAGGCTGTGGGCGTGGGCGACAGGCAAGAGCCATTACTTTGTGCGCAAGCAAGAT 2438  
y 2101 AACCGGTGACGAGAAATGATGGAATGCTTCAAGCAGACAAACGGCATGCTCTCTTGTG 2160  
b 2439 GGAGGACAGGCTGGATCCATGATGCTGCTACAGTATGACGAGCCCTGTGCGCAGCCCTTACC 2498

QY 2161 CAAGAATGCACAGTCCCATGTGAGAGAGTGCACCTTCACTTGTGTTCCAGTTTACG 2220  
Db 2499 CAGGCTGCCAGATCCCTGCCAGGATGATGTCAATTCACAGCTGGTCCAGTTTCT 2558  
QY 2221 CCTGTCTCCAGAACTGTGAGGCCAACAAAAGTGAAGGCGGACAGCTCAGGGAAGC 2280  
Db 2559 TCATGCAATGAGAGCTGTGTGCAAGTGAAGCCAGAAAGCGCACTTGTGTGAAAGT 2618  
QY 2281 AGAAGAGGAGAAATGCCAGGATCTGACCTTTTACCTCTAGTGAGAGCAGAACTATGT 2340  
Db 2619 AAAAGAGGAAATGTAATAATTTCCATTTGATATCCCTGATTCAGACTCAGTATTGT 2678  
QY 2341 CCTTGTGATGAATTTATATCCCAACCTTATGGAATGTGTCAAGTTGCAATTTCTTCAGAA 2400  
Db 2679 CCTTGTGACAAATATAATGCACAACTGTGGGAACTGTGTCAAGCTGTATTTTACCAGAG 2738  
QY 2401 GGCAGAGGAGGCTCACCGAGACTGCGGTTCAAGCAGACAGCAAGAAATGTGGAGAA 2460  
Db 2739 GGAAGTGGAAATGTGTGCGGAAATGAAGTACAAGGAGACATCAAGGAATCGGACAA 2798  
QY 2461 GGCCTGGCTTTCGAGCAGTAGCTGTCTGTATAAAAAATGGAAGACCTGTGTGACCCCTCC 2520  
Db 2799 GGATATGTTACCAAGCAATGGCATGCTACGATCAAAAATGGCAGGCTTGTGGAACATCT 2858  
QY 2521 TTCTGACAGAGCTCTGTGTACATTCAGAAAAATGTGTCTATTCCTGCCCAATTTGATTGC 2580  
Db 2859 AGATGTAAACAGCCATGTTTACATTTGAGGAGGCTGCATCATCCCTGCCCTCAGACTGC 2918  
QY 2581 AAGTTAAGCAATTTGTGTAGTTGGGGTCTTGTGAGTTCATCTTTGTGGAATTTGGAGTGAGA 2640  
Db 2919 AAGCTCAGTGAAGTGGTCCAACTGGTGGCTGCGAGTCCCTGTGGAGTGTGTGAAG 2978  
QY 2641 ATTGATCCAAATTTGGCTTAAAGAAAAACCTTACAATGAGGAGACGACATGTCCCAACTG 2700  
Db 2979 GTTCGTTCTAAATGGCTGCGTGAATAACCATATAATGGAAGAAAGGCTTGTCCCAACTG 3038  
QY 2701 GATC-----TCAGAAATCAGGTACATGAGGAGTCCCATGTTTACAGTGAAGTCAATCAG 2754  
Db 3039 GACCATGTCAACAGGACAGGTGTATGAGTTGTCCCATGCCACAGTGAATGCACCCAG 3098  
QY 2755 TATTCTGGGTGTGAGAACACTGGTCTTCAATGCAAAATCAACAATGAGCTGAGGTCCCTG 2814  
Db 3099 TACCTATGGGTCAACAGAGCCCTGGAGCATCTGCAAGTGACCTTTGTGAATATATCGGGAG 3158  
QY 2815 CGCTGTGGAGGAGAACACAACTATGGAATAATCAGATGTGT---GAATCTGGGATGGT 2871  
Db 3159 AACTGTGGAGGCGGTGCAACCCGAAAGTGAAGATGCAATGAGATGAGTGCAGAAATACAGAGATGGC 3218  
QY 2872 GAAGGTGGAGCAGTGGATAGCAACCTGTGCAACCCAGGATGAAATTTCCCCAGAAACCCAG 2931  
Db 3219 CCTTCTGAACATGTAGAGGATTTACCTCTGTGACCCAGAGAGATGCCCTGGGCTCTAGA 3278  
QY 2932 TCCTGTTCTTATGTGTCCCAATGAGTGTGTCTATGCTGAGTGGGACCTTTGGAGCAAA 2991  
Db 3279 GTGTGCAAAATTACCATGCCCTGGAGGACTGTGTGATATCTGAATGGGGTCCATGACCCAA 3338  
QY 2992 TGCCCAAGTCAATCGCATCCCAACAAT---CGAGAGAAAGAACTCGCCACCTGCTTAAGA 3048  
Db 3339 TGTGTTTGTCTTCATCAAGAGTTCGCGCAAGGTTCCGCAAGGTCAGTGTATCCCATCAGACAA 3398  
QY 3049 CCATCACTGAACCTCAAGAGCTTGTGTGAAGATCTCAGGTGAGCCCTGCTCTCTGAAT 3108  
Db 3399 CCAGCTGATGAAGAAAGATCTTGGCCCTAATGCTGTTGAGAAAGAACCTGTGAACCTGAAC 3458  
QY 3109 GAAATGCTTCCAGTTCCAGTACAATCTAACAGGTGGAGCACATGCCAGCTCAGTGA 3168  
Db 3459 AAAAATGCTTACCACTATGATTAATGTAACAGCTGGAGTACATGTCACTGAGTGAG 3518  
QY 3169 AACGCACTGTGTGTAAGGCGTCAAGACCCGCTGCTAAGCTGTGTGTGAGTGTGTCAGTGTGTC 3228  
Db 3519 AAGGAGTGTGTGGAATGGAATAAAAAACAAGGATGTTGGATTGTGTTCGAAGTGTGTC 3578

3229 AAGCCAGTCAGCATGGCAATGTGACGACATAATTTGGAGAAAGCCCCAGAGAAATGAGC 3288  
3579 AAGTCAGTCGACCTGAATATTGTGAAGCCCTTGGCTTGGAGAAAGAACTGCGCAGATGAAC 3638  
3289 ATTCCCTGCTTGGTGAATCGGTGCTCACTGTCACTGTCAAGGTTGAGCGGCTTGACA 3348  
3639 ACGTCCCTGCAATGGTGAATCCCTGTGAATGTCAAGTCTTCTGATGGTCTCTCTGGTCA 3698  
3349 GAGTGTTCACAGACCTGTGCGCATGGAGTCGATGAGCCGAGCTCGATTATCAATTATG 3408  
3699 GAATGTTCTCAACATGTGCGCTCACAGGAAATGATCCGAGAGCAACAGTGCACCCAG 3758  
3409 CCAACCCAGAGAGAGAGCGCGATGCCCCACAGAGCTTACCCAGGAGAAACCTGCCCA 3468  
3759 CCGTTTCAAGGTGATGGAAGACCATGCGCTTCCCTGATGAGCAGCTCCAAACCTGCCCA 3818  
3469 GTGACCCCTGCTACAGCTGGTCTTGGCAACTGGTCTCATGTAATTTGGAGGTTGGA 3528  
3819 GTGAGGCTTGTATCGTGGCAATATGGCCAGTGGTCTCCATGCCAAGTGCAGGAGGCC 3878  
3529 GACTGTGGGAAGAGTTCAGATCCGAGCGCTTTCCTGCAATGTTCCACAGTGGTCAATA 3588  
3879 CAGTGTGGAGAGGACCAAGCAAGCAACATTTCTGTGTAGTGAATGATGGTCAAGCT 3938  
3589 TCTCATGCTGAGCGTGTGAGGATGCTGTGAGGAAATGCCCTT----- 3638  
3839 GATGATTTACGAAAGTGGTGGATGAGAAATTTCTGCTGCAATGAACTCATATAGAT 3998  
3639 -----TCAGGACAGCATCCTGAAGCAGCTGTGTGTGCTGCTTCCCGAGAGACTGCCAT 3693  
3999 GGTAAATAAAATATGTTCTGGAGAAATCCTGCAGCCAGCGCTTGCAGGTGATTTAT 4058  
3694 TTACAGAAATGTCAGAGTGAGACATGTGATTAACCTGCATTTGATGAGAAAGCTTT 3753  
4059 TTGAAGGACTGGTCTTCTGAGGCTGTGTGAGTGCAGTGCAGTGTGATGAGGATCTA 4118  
3754 GAGACTGTGGGCCCGCCAGCTAGATCAAGACATTTTATAATTCAGTCTTTTGAAGAACCA 3813  
4119 GGCTTTGGTGAATACAGTTCAGATCCAGACCGGTGATTATACAGAACTAGAGATCAG 4178  
3814 GACAGTGGCCCCAACAGGTTCTAGAAACAGCGCTTGTACAGAGGCAATGTTATCAC 3873  
4179 CATCTGTGCCAGAGCAGATGTTAGAAACAAATCATGTTATGATGACAGTCTGTATGAA 4238  
3874 TACATATGGAAGCAAGTCTTTGAACAATAACGAACTGTATGTCGCCAGCGTCTCA 3933  
4239 TATAAATGATGGCCAGTGTGGAAGGGCTCTTCCGACAGTGTGGTGTCAAGGTCA 4298  
3934 GATCGGTTAATGTCACAGGAGG 3956  
4299 GATGGTATAAATGTAACAGATGG 4321

RESULT 15  
ADB62804  
ID ADB62804 standard; cDNA; 3053 BP.  
XX AC ADB62804;  
XX AC  
XX AC  
DT 04-DEC-2003 (first entry)  
XX Human cDNA encoding clone NT2RP70056690.

XX Human; ss; gene; pharmaceutical; diagnostic; gene therapy;  
XX tissue regeneration; cell regeneration; membrane protein;  
XX signal transduction-related protein; transcription-related protein;  
XX osteoporosis; neurological disease; cancer; tumour.  
XX Homo sapiens.

XX Key Location/Qualifiers  
XX FH 253..3053  
XX CDS /\*tag= a

/partial  
/product= "Clone NT2RP70056690 protein"  
EP1308459-A2.  
07-MAY-2003.  
28-MAR-2002; 2002EP-00007401.  
05-NOV-2001; 2001JP-00379298.  
25-JAN-2002; 2002US-00350978.  
(HELI-) HELIX RES INST.  
(REAS-) RES ASSOC BIOTECHNOLOGY.  
Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;  
Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;  
Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;  
WPI; 2003-450961/43.  
P-PSDB; ADB64774.  
New polynucleotides and polypeptides, useful for developing a diagnostic  
marker or medicines for regulation of their expression and activity, or  
as targets of gene therapy.  
Claim 1; Page; 22pp; English.  
The invention discloses a polynucleotide comprising a sequence selected  
from 1970 fully defined nucleotide sequences which encode novel  
polypeptides. Also claimed is a polypeptide encoded by the polynucleotide  
or its partial peptide, an antibody binding to the polypeptide or peptide  
of the polynucleotide, immunologically assaying the polypeptide or  
peptide of the polynucleotide by contacting the polypeptide or peptide  
with the antibody of the encoded protein, and observing the binding  
between the two, a transformant carrying the polynucleotide in an  
expressible manner and an antisense polynucleotide. The oligonucleotide  
is useful as a primer for synthesising the polynucleotide, or as a probe  
for detecting the polynucleotide. The polynucleotides and encoded  
proteins are useful as pharmaceutical agents and many disease-related  
genes may be included in them, for developing a diagnostic marker or  
medicines for regulation of their expression and activity, or as targets  
of gene therapy. The genes are involved in tissue and/or cell  
regeneration. Membrane proteins, signal transduction-related genes,  
transcription-related proteins, disease-related proteins and genes  
encoding them can be used as indicators for diseases (e.g. osteoporosis,  
neurological diseases, cancer, tumours). The cDNA may be used to regulate  
the activity or expression of the encoded protein to treat diseases. The  
sequence presented is a cDNA of the invention. Note: Some of the sequence  
data for this patent is not represented in the printed specification, but  
is based on sequence information supplied by the European Patent Office.  
Sequence 3053 BP; 780 A; 783 C; 891 G; 599 T; 0 U; 0 Other;  
Query Match 13.2%; Score 580.6; DB 9; Length 3053;  
Best Local Similarity 56.4%; Pred. No. 2.7e-172;  
Matches 1309; Conservative 0; Mismatches 934; Indels 78; Gaps 9;  
QY 10 TGCATTCAAGAGCTCAACCGAACTGTGTTGCAATGAAATATATCGAACACTTTGCCCTT 69  
DB 730 TGCATCCAGAAAGACAAGACATTCCTCGGAGGATATCATCTGTGAGTACTTTGAGCCC 789  
QY 70 CAGCCTCTCAGAACAGAGCTTGCCTCATTCCTTGTCCCGGATTTGTAGTATCTGAG 129  
DB 790 AAGCCTCTCTGGAGCAGGCTTGCCTCATTCCTTGCAGCAAGATTGCATCTGTCTGAA 849  
QY 130 TTCTTACCATGTCCAACTGTAGCAAGGATGTGGAGAAATTTGCAGCATAGAACTCGC 189  
DB 850 TTTTCTGCTGCTCGAATGCTCCAGACCTTGGCAGCGGCTCCAGCCGCGCT 909  
QY 190 GCGGTCATAGCTCCCCCTCTCTTTGGTGGTTTGCATGTCCAAATCTGATGATCAAGA 249  
DB 910 CATGTGGTGGCGCTCCGAGCTTCGAGGCTCTGGCTGTCCAAACCTGACGAGTTCAG 969

250	GCCTGTGATGCTCCCATTTCTCTGTCTCTTTGGGAAAGAGGAATATATCATTTAGCCTTAAG	309
ib		
970	GTGTG-----CCAATCCAGTCCATCGAGGCGGAGGAGCTCAGGTACAGCCTGCAT	1020
ib		
310	GTGTGACCATGGAGTAAATGCGACACTGCTCTTAAAGAAATTAATCCAAAGCGGAAGA	369
ib		
1021	GTGGGGCCCTGGAGCACCTGCTCAATGCCCCACTCCCGACAGTAAAGACAAGCAAGGAGA	1080
ib		
370	ACTGTTCTGGATTTTAA-----CTCTGATTCAAATGAG	402
ib		
1081	CGCGGGAAGATAAAGAAACGGGAAAGAGACCGCAGCAAGGAGTAAAGGATCCGAGAGCC	1140
ib		
403	CGAGTCACCTTTAAACATCAAAAGTTACA-----AAGCACATCATCATTCGAAAGTCT	453
ib		
1141	CGCGAGCTTTATTAAGAAAAGAGAAACAGAAACAGCAGACAGACAGAGAACAATAT	1200
ib		
454	TGGGCAATAGAGATAGTTATCAAACCGGCAGGTTTCGTGTACAAGAGTGTATGGACAA	513
ib		
1201	TGGGACATCCAGATTGGATATCAGACACAGAGAGTTATGTGCATTAAACAGACGGGGAAA	1260
ib		
514	AATGTATGCTTAAGCCCTTTGGCCTTCAAGATTCTTCCCATTTGACTCTTCAGTCTCTGCATC	573
ib		
1261	GCTGCTGATTTAAGCTTTTGGCCAGCAGAGAGACTTCCAATGACCTTCAGTCTCTGTGTG	1320
ib		
574	ATGCCCAAAGACTGTGAACCTCCAGTGGTCTCTCTGGAGCCCTGTCTCAAAGACATGC	633
ib		
1321	ATCACCAAAGAGTGCAGGTTTCCGAGTGGTCAGAGTGGAGCCCTGTCTCAAAAACATGC	1380
ib		
634	CGTTCAGGAGTCTCTTTGCCAGGATTTAGAGCAGGAGCCGGAACGTGAAGCACATGGCT	693
ib		
1381	CATGACATGGTGTCCCTCGAGGCATCTGTGTAAAGACAGAAACCATCAGGCAGTTCCCT	1440
ib		
694	ATTGGAGTGGAAAGAGGTCTCTGAACTTCTTGAAAGAGAGCCCTGCATTTGTGAAGA	753
ib		
1441	ATTGGCAGTGAAGAGAGGTCTCAGAAATTTGAAGAAAAAGAACCTGTTTGTCTCAAGA	1500
ib		
754	GAACITCTCAGCAATGCCAGGTATTCCTGGAGAACTTCTGAATGGAAGAAATGCCAA	813
ib		
1501	GATGGATTGTCCCTGTGCCAGTATGGCTGGAGAACTACAGATGGACTGAGTGCCCT	1560
ib		
814	GTCTCTCTCCTCCTCGAGCAGCAGGATCCCACTGGCATGTGACGGGACCCGCTGTGTGC	873
ib		
1561	GTGGACCTTTGCTCAGTCAGCAGGACAAGAGCGCGGCAACCAGACGGCCCTCTGTGGA	1620
ib		
874	GTGGGATCCAGACCCGGGAGGTACTGTGCCAG-----AGCGTACCAGCA	921
ib		
1621	GGGGGCATCCAGACCCGAGGAGGTACTCGTGCAGGCCCAACGAACCTCTCTCACAA	1680
ib		
922	GCTGCCGCACTCAGGCGCCAGGAAGTCTCTAGACCTGTGGAAAAGGCATTATGTGTGGGA	981
ib		
1681	TTAAGTACCACAAGAAACAAAGAAGCCTCAAAGCCAAATGGACTTAAATATATGCACCTGA	1740
ib		
982	CCCGCCCGCTTGCCCTCTCAGCTCTGCAATATCCCTTGCTCTACGGAAGTGCATAGTATCT	1041
ib		
1741	CCATCCCTTAATACTACAGCTGTGCCCATTTCTTGTCCAACTGAATGTGAAGTTTCA	1800
ib		
1042	TCCTGTGCACTCGGGCCCTGTGCATCCATGAAACTGTGTAAGAAAGGCATTATGTGTGGGA	1101
ib		
1801	CCTTGTGAGCTTGGGACCTTGTAATTATGAAAACCTGTAATGATCAGCAAGGGAAGAAA	1860
ib		
1102	GGATTTAGAACGAGGCAGCG-----CCATGTCTCATTGGAACTTACAGGGCCCTGCA	1152
ib		
1861	GGCTTCAAACTGAGGAAGCGGCGCATTTACCAATGAGCCCACTGAGGGCTCTGGGGTAAAC	1920
ib		
1153	GGGCATGCCCCATTTGGTGGAGTCTGTCTCCCTGTGTGAGGATCCAATGTGCTACCGATGG	1212
ib		
1921	GGAACTGCCCTCACTTACTGGAGCCATTCCTCTGGAAGAGCTGCCCTTTATGACTGG	1980
ib		
1213	CTGGCA---TCAGAAAGGATCTGTTTCCCTGATCATGGAA---AATGTGCCCTGGACAT	1266
ib		
1981	AAAGCAGTGAAGCTGGGAAACTCGAGCAGCAGATAACGGAAAGAGGTGTGCTCCAGGCAGC	2040
ib		

1267	QY	CGTATTCTGAAGCCCGTCTGCGCAGAAATGACCGGGAGAAGATGTATCAGGGAGTCTTTGC	1322
2041	DB	CAAGTTCAAGAGGTTGTGTGTCATCAACACGTGATGGAGAAGAAGTTGACACAGCAGTGTGC	2100
1327	QY	CCAGTTCGCCCTCTCTCAGAGGAAGTCTTGTGAAATCCCTGTCCGGAATGGAAGTGTGTG	1386
2101	DB	AGAGATGCCATCTTCCCATCCTGTGGCTGTGTATGCCCCCATGCCCGAAGACTGTGTG	2160
1387	QY	CTGAGCGAGTGGACGAGTGGTTCATCTCTGTTCCAGTCTCTGTTCAATAAAACTCAGAT	1446
2161	DB	CTCAGCACATGGTCTACGTGGTCCCTCTCTCACACACCTCTCAGGAAAAACGACAGAA	2220
1447	QY	GGGAAACAGACACAGGTCAAGAACTACTCTGGCACTGGCTGGG---GAAGGTGGAAAGCCA	1503
2221	DB	GGGAAACAGATACGACACGATCCATCTCTGGCCTATGCGGGTGAAGAGTGGAAATTCG	2280
1504	QY	TGTTCCCTTAGTCAGCTCTCCNAGAGCATGTTTGTGTAAATGACCAATTCCTGTATGCGAG	1563
2281	DB	TGTTCCAAATAGCAGTCTCTTGAAGAAGTACGAAGCTGTAATGAGCATCTCTTGCACTG	2340
1564	QY	CTTCACTGGGAGACATCGCTCTGGGGCCCTTGTCTGAGGACACATTTGTTAACTGCGCCTT	1623
2341	DB	TACCACTGGCAACTGGTCCCTGGGGCCAGTGCAATTGAGGACACCTCAGTATCGTCTTTC	2400
1624	QY	AATGCAACCATTTGGCTGGAAATGGAGAAGCAACGTGTGGTGTAGGCCATTCAGACTCGAGA	1683
2401	DB	AACACAACACGACTTGAATGGGAGGCGCTCTCTCTCTGTGGCATGACAGCAAGAAAA	2460
1684	QY	GTCTTCTGTCTCAAGAGTCACTGGGACCAAGTAATGACCAAAAGATGTCACAGATTTCTACT	1743
2461	DB	GTCACTGTGTGGAGTCAATGTGGGCCAAGTGGGACCCAAAAATGTCCTGAAGCCCTT	2520
1744	QY	CGACTGAAACTGTGCGCCCTGTTTTCTCCCATGCAAAAAAGACTGTATTTGTGACTGCT	1803
2521	DB	CGACCTGAAACTGTAAAGGCTTGTCTGCTTCCCTGTGTAAGAGGACTGTATTTGACCCCA	2580
1804	QY	TTCAAGTGAAGTGGACACCTCTGCCCAAGGATGTGCCAAGCAGGAAATGCCACAGT---	1860
2581	DB	TATAGTGACTGGACATCATGCCCTCTTCTGTGTAAAGAGGGGACTCCAGTATCAGGAAG	2640
1861	QY	CAGTCTCGATACAGAAATCATCTCAAGAAGCAGCCCATGGAGGCCAGAAATGCCAGAT	1920
2641	DB	CAGTCTAGGCATCGGGTCTCATTTCACTAGCTGCCAGCCAAACGGGGCGGAGACTGCACAGAT	2700
1921	QY	ACCTTATATGAGGAGAGAGTGTGAAGATGTTTCCTTGTGTCCTGTATATCGTGGGAAG	1980
2701	DB	CCCTCTATGAAGAGAAAGCCCTGTGAGGCACCTCAAGCGTGCCAAAGCTACAGTGGGAAG	2760
1981	QY	CCACAGAAATGGAGCCCTTGCATCTTAGTGCCAGAGTCTGTCTGGCAGGGAATAACGGGC	2040
2761	DB	ACTCACAAATGGCGCAGATGCCCAATTAGTCTCCCTTGAGCGTGTGCAACAGACAGCCCTGA	2820
2041	QY	AGCAGTGAAGCCCTGTGGAAAGGGTTTACAAAACAGAGCTGTCTCATGTCATCTCTGATGAC	2100
2821	DB	GCACAGGAAGGCTGTGGGCCCTGGGCGACAGGCAGAGCCATTTACTTGTGCGAAGCAAGAT	2880
2101	QY	AACCGTCTAGCAAAATGATGGAAATGCCTTAAGCAGACAAAAACGGCATGCTCTCTTGTG	2160
2881	DB	GGAGGACAGGCTGGAATCCATGAGTGCCTACAGTATGCAGGCGCCCTGTGCGAGCCCTTACC	2940
2161	QY	CAAGAAATGCACAGTCCCATGTCTCGAGAAGACTGCACCTTCCACTGCTTGGTCCAAGTTTACG	2220
2941	DB	CAGGCTGCCAGATCCCTCTGCCAGATGACTGTCAATTTGACCAAGCTGGTCCAGTTTCT	3000
2221	QY	CCCTGTCTCAAGCACTGTGAAGCCCAAAAAAGTAVGGCGCG	2261
3001	DB	TCAATGAATGGGACTGTGGTGCAGTTAGGACACAGAAAGCG	3041

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

M nucleic - nucleic search, using sw model

run on: February 23, 2004, 11:46:56 ; Search time 10047.4 seconds  
(without alignments)  
13071.481 Million cell updates/sec

Title: US-10-022-710-1

Perfect score: 4398

Sequence: 1 atgttgctgctgattcagaa.....atggagacttagacatgtaa 4398

Scoring table:

IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_hic:\*

9: gb\_estl:\*

10: gb\_est2:\*

11: gb\_hic:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: em\_gss\_hum:\*

18: em\_gss\_inv:\*

19: em\_gss\_pln:\*

20: em\_gss\_vrt:\*

21: em\_gss\_fun:\*

22: em\_gss\_man:\*

23: em\_gss\_mus:\*

24: em\_gss\_pro:\*

25: em\_gss\_rod:\*

26: em\_gss\_pbg:\*

27: em\_gss\_vrl:\*

28: gb\_gssl:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1963.4	44.6	3897	11	AK051714 Mus muscu
2	1908.8	43.4	3242	11	AK083798 Mus muscu
3	866.4	15.2	2266	11	AK047508 Mus muscu
4	621.8	14.1	676	10	BE299433 601118764

5	600.4	13.7	760	10	BE295719
6	585.2	12.9	772	14	CF729699
7	543.4	12.4	757	14	CB723142
8	542.6	12.3	820	13	BU703731
9	506.8	11.5	717	13	BY735628
10	485.8	11.0	683	10	BB451562
11	472.6	10.7	600	12	BI988413
12	451.6	10.3	565	9	AV590588
13	447.2	10.2	713	12	BM414542
14	446	10.1	551	10	BE899999
15	422.4	9.6	600	12	BI981468
16	421.6	9.6	600	13	BU251229
17	404.8	9.2	491	10	BF605229
18	354.2	8.1	670	10	BB616046
19	310.6	7.1	693	13	BU313848
20	288.2	6.6	531	9	AL713171
21	283.4	6.4	676	14	CF531936
22	272.6	6.2	535	10	AW787130
23	243.2	5.5	760	13	BU262534
24	241	5.5	735	13	BU283850
25	239.4	5.4	750	13	BU308368
26	239.4	5.4	793	13	BU200855
27	223	5.1	658	13	BU267027
28	214	4.9	944	13	BU366211
29	212.6	4.8	316	10	BA452816
30	201	4.6	828	13	AQ531824
31	197	4.5	738	10	BU451181
32	196.2	4.5	556	10	BE232461
33	195.2	4.4	649	14	CF362876
34	194.6	4.4	801	13	BU318139
35	194.6	4.4	523	12	BI340182
36	189.4	4.3	503	10	BF441768
37	189.2	4.3	673	28	AQ737047
38	184.8	4.2	531	28	AQ737047
39	181.4	4.1	677	10	BA463698
40	179.4	4.1	529	12	BI340184
41	178.2	4.1	928	13	BU267473
42	168	3.8	695	13	BU451889
43	167.6	3.8	583	13	BU100582
44	166	3.8	438	10	BF441422
45	165.6	3.8	475	10	BF441417

## ALIGNMENTS

RESULT 1  
AK051714  
LOCUS  
DEFINITION  
Mus musculus 12 days embryo spinal ganglion cDNA, RIKEN full-length enriched library, clone: D130067I03 product: weakly similar to HYPOTHETICAL PROTEIN DJ1110N13.1 IN CHROMOSOME 7 (FRAGMENT) [Homo sapiens], full insert sequence.

ACCESSION  
AK051714

VERSION  
AK051714.1

KEYWORDS  
HTC; CAP trapper.

SOURCE  
Mus musculus (house mouse)

ORGANISM  
Mus musculus

REFERENCE  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS  
Carninci, P. and Hayashizaki, Y.

JOURNAL  
Meth. Enzymol. 303, 19-44 (1999)

MEDLINE  
9279253

PUBMED  
10349636

REFERENCE  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

TITLE  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL  
Genome Res. 10 (10), 1617-1630 (2000)

MEDLINE  
20499374

```

11042159
3
PUBMED
REFERENCE
AUTHORS
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,
Sunii, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, K., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanishi, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--394-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
MEDLINE
PUBMED
11076861
AUTHORS
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 585-590 (2001)
5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6
(bases 1 to 3897)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saigo, R., Saichon, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akanira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (18-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan [E-mail: genome-res@gsc.riken.go.jp].
URL: http://genome.gsc.riken.go.jp/ Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/.
Location/Qualifiers
1. 3897
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:D130067103"
/db_xref="MGI:2420048"
/db_xref="caxon:10090"
/clone="D130067103"
/tissue_type="spinal ganglion"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="12 days embryo"
174..3497
/note="unnamed protein product; putative
weakly similar to HYPOPHYSICAL PROTEIN D1110N13.1 IN
CHROMOSOME 7 (FRAGMENT) [Homo sapiens] (SPTP:O43376,
evidence: FASTY, 53.6%ID, 99.4%length, match=540)"
/codon_start=1
/protein_id="BAC34732.1"

```

```

/db_xref="GI:26342140"
/translation="MFLRSDLAVTHWYSRNRKFLVLSLLSOAAHLEGRKDNQFLW
KTGPMWGCAGDCGPGQSRVAVCFHLEGWTSFMSNCDSESSQPPKESCFVCDMDHSD
LFOVEYDWRHRCILVPGAGEPRVAVCVTAQGLQHRVCLQKLRNTWVNEICE
HPAPOPTEOACILIPCRDQCVSEFSPWSTCPGCGKLOHRTVETATAPLYGLOCE
NLTSERACEAPVSPKLEYSLSKLVGPMKSLPHLKVLDLSGRNIQDFSSDSNQ
VTLTHQSKAHHSQKQGVVIGFQTRQVWCTRSDGRNALSLCVRDSFPFLTVQPCWMP
XDCTSSPWSWPCSKTCRSQTLSPGVRSSRNKVIHAGGSGTQTRVYCAQSLPAGI
EFLQPCRYSPWTEWKEQVSLLEQHPMHTGPICGGGTQTRVYCAQSLPAGI
ASRTKEVSRPVERTLCIGPAPASQOLCNVPCMDCIVSSWSTWGPVCFVFNCHDQGGK
GPRVQRHLMESTGPMCRPHLAEVPCEDPMCHRWLASEGICIAHDGKCGLGHRIIL
KAVCONSGEVSGLCLPVPPEPMACEIPCRMDVSVSWTVSSQSCSNKNSDG
KQTRSRILALAGCGGKTCSSQLEQYRLNDHSCOTLYWTSAMGSCSENTVATLAL
VATFSEWTPCPRPCPGFNTIIQKSYRIIIQEAANGQCEPDILFEERECEDKDCCLV
RWKQVQSECLVPEDSIRQGTSTSEACGKGLQTRAVSCISDDNQSAEMTECLQMNG
MPVLVQECTIPCRDQCVSEFSPWSTCPGCGKLOHRTVETATAPLYGLOCE
LVNEMPCPDFTMSHPYGNWASACILIPGKRDQOGLWQVQDKECEGVRFRFRAIAGS
NINRGPVDPSPFNSSGVYQAEACVPCPFDCILSDWSWSSGSSSGGIGVIRSKWLKE
KPYSGGRPCPKLDLKNQFCHLSISSGATLFLKILLMDMSFLFAVVCFCYCFQSLFL
ICVFQRLVHQVFRDLRCLGA"
polya_signal
3880..3885
/note="putative"
polya_site
3897
/note="putative"
ORIGIN
Query Match 44.6%; Score 1963.4; DB 11; Length 3897;
Best Local Similarity 82.9%; Pred. No. 0;
Matches 2253; Conservative 0; Mismatches 461; Indels 3; Gaps 1;
QY 4 GTGCCTGCATTCAAGCTCAACCGAACTGTGGTTGCAATGAAATGAAATGCGAACACTTT 63
DB 600 GTGCCTGTCTTCAGAGCTCAATAGAACCATGGTGTCCATGAGATCTGTGAACACTTT 659
QY 64 GCGCTTCAGCTCTTACAGACAGAGCTTGCCTCATCTCTTGTCCCGGGATGTGTAGTA 123
DB 660 GCTCCTCAGCCCCCAGAGCAAGCTGCTCTCATCTTCCCTCCCGAGGACTGGTGTGTA 719
QY 124 TCTGAGTTCTTACCATGTGTCCAACTGTAGCAAGGATGTGGAGAAATTCAGCATAGA 183
DB 720 TCTGAGTTCTTCCCATGTGTCCAACTGTCCCGAGAGATGTGGAGAAATTCAGCATAGA 779
QY 184 ACTCGCGGGTCAATAGCTCCCCCTCTCTTTGGTGGTTTGCATGTGCAATCTGCAATGAG 243
DB 780 ACTCGTGTGCCATCGCGCCCCCTCTGTATGAGAGTCTGCAAGTGTCCAAATCTCACGAG 839
QY 244 TCAAGAGCTGTGATGCTCCCATTTCTCTCTTGGGAGAGGATATACATTTAGC 303
DB 840 TCCAGAGCTCGGAGGCTCCAGTTTCTCTCTTGGGAGAGGATATATTCCTTCAGC 899
QY 304 CTTAAGGTTGACCATGGAGTAATTCAGACTGCTCTTCTTAAAGAAATTAATCCAGC 363
DB 900 CTTAAGGTTGACCATGGAGTAATTCAGACTGCTCTTCTTAAAGAAATTCAGC 959
QY 364 GGAAGACTGTCTGGATTTTAACTCTGATTCAAATGAGCGAGTCACTTTAAACATCAA 423
DB 960 GGAAGAAATATCCAGGATTTTAGCTCTGATTCAAACGAGAGTCACTTAAACATCAG 1019
QY 424 AGTTTACAAAGACATCATCATTCGAAAGTCTTGGGCAATAGAGATAGGTTATCAAAACCGG 483
DB 1020 AGTTTACAAAGACATCATCATTCGAAAGTCTTGGGCAATAGAGATAGGTTATCAAAACCGG 1079
QY 484 CAGGTTTCGTGTACAAAGATGATGCAAAATGCTATGTTAAGCCTTGGCTTCAAGAT 543
DB 1080 CAGGTTTCGTGTACCAAGATGATGCAAAATGCTATGTTAAGCCTTGGCTTCAAGAT 1139
QY 544 TCTTCCCATTCAGCTGTTCCAGTCTTCATGCTCCCAAGACTGTGAAACCTCCAGTGG 603
DB 1140 TCTTCCCATTCAGCTGTTCCAGTCTTCATGCTCCCAAGACTGTGAAACCTCCAGTGG 1199
QY 604 TCTTCCCATTCAGCTGTTCCAGTCTTCATGCTCCCAAGACTGTGAAACCTCCAGTGG 663

```

1200	TC	TC	CAATGGAGCCCTCTGTTCCAAAGACATGCGGTTCTGGGACTTTTATCACTCGAGGTAGA	1259
664	AG	CAGGAGCCGGAACTGAAAGCACATGGCTATTGGAGGTGGAAAGGAGTGTCTCTGAACTTT	723	
1260	AG	CAGGAGCCGGAACTGTAAGCATATTGCTATTGGAGGTGGCAGGAATGCCAGAACTT	1319	
724	CT	TGAGAAAGAGCGCTGCATTTGTTGAAGGAGAACTTCTGACGACAAATGTCACAGGATATCC	783	
1320	CT	TGAGAAAGAGACCTGCATTTGCTTGAAGGGGAATTTCTGAACCATGTCTCCAGGATATCC	1379	
784	TG	GAGAACTTCTGAATGGAAGAATGCAAGTCTCTCTCTCTCTCTCGACGACGAGATCCC	843	
1380	TG	GAGAACTTCTGAATGGAAGGATGCCNAGTCTCTCTCTCTCTGAGGACACGATCCA	1439	
844	CA	TCTGGCATGTGACGGAGACCGGTGTGTGGCGGTGGATCCAGACCCGGGAGGTGTACTGT	903	
1440	CT	TTTGGCATGAGACAGGACCAATCTGTGTGTGGGATCCAGACCCGGGAGGTGTACTGT	1499	
904	GCC	CAGAGCTTACCAGAGCTGCCGCACTGAGGCGCAAGGAAGTCTCTTAGACCTTGTGGAA	963	
1500	GCC	CAGAGCTTACCAGAACCAATTGCATCAAGGACCAAGGAAGTCTCCAGGCCGTGGAA	1559	
964	AAG	GCAATTAATGTGGGACCGCCCGGTGGCCCTCTCAGCTCTGGAATATCCCTTGCTCT	1023	
1560	AG	GACACTATGCTTGGGACCTGCCCTCTCAGCTTTCGAATGTCCCTTGCTCTG	1619	
1024	AC	GNCTCAGTAGTATCTTCTCTGCTCAGCTCGGCCCTGTGCATCCATGAATACTGTCTAT	1083	
1620	AT	GGACTGCATAGTATCTTCTCTGTCAACCTGGGCCCATGTGTGTTTTGAGAACTGCCAT	1679	
1084	GA	ACCTCAGGGGAAAAAGGATTTAGAACGAGGCGAGCCCATGTCTCATGGAACTTACA	1143	
1680	GAT	CTCAGGGGAGAAAGGCTTTAGATGAGACAGCGCCATGTACTCATGGAGTCTACA	1739	
1144	GG	CCCTGACGGGCATTTGCCCTCATTTGGTGGAGTCTGTTCTCTGTGAGATCCCAATGTC	1203	
1740	GG	CCCATTGGGGCGTGCCTCATCTGGCAGAATCTGTACCTTGTGAGATCCCAATGTGT	1799	
1204	TAC	CGATGGCTGGCATCAGAAGGATCTGTTTTCCCTGATCATGSAATAATGTGGCTGGGA	1263	
1800	CAT	CGCTGGCTCTCGAAGGGATCTGTAATGCTGATCATGGGAATGTGGCTGGGA	1859	
1264	CAT	CGTATTTGAAAGCCGTCTGCCAGATGACCGGGAGAGATGTATCAGGGAGTCTT	1323	
1860	CAT	CGCATTTCTAAGGCTCTCTGTCTCAGAAATGACGGGAGAGAAAGTATCAGGGGCTCTC	1919	
1324	TG	CCAGTTCCCTCTCTCTCAGAGAGAACTTTGTGAAATTTCCCTGCCGAATGGACTGT	1383	
1920	TG	TCCAGTGCCTTCTCTCTCTGNAAGAAATGGCTTTGAAATTTCTTGCCGAATGGACTGT	1979	
1384	GT	CTGAGCGAGTGGACGAGTGGTCAATCTGTGTTCCAGTCCCTGTTCAAAATAAAACTCA	1443	
1980	GT	AGTGAAGTGGACAGTGTGGTCAATCTGTCTCAGTCTCTGTTTCAAAATAAAACTCA	2039	
1444	GAT	GGAAACAGACACCAAGTCAAGAACTATCTCTGGCACTGGCTGGGGAAGTGGAAAGCCA	1503	
2040	GAT	GGGAAACAAACCAAGTCAAGTCTTATCTCTGCTTGGCTGGGAGAGTGGAAAGACC	2099	
1504	TGT	CCCCCTAGTCAGGCTCTCCAAGAGCATCGTTTGTGTAATGACCATTCCTGTATGCGAG	1563	
2100	TG	CCCTAGTAGTCAGGAACTCCAAGAGTACCGCTATGCAATGATCATTTCTCTGTACACAG	2159	
1564	CTT	CACCTGGGAGACATCGCTTGGGCCCTTGTCTGAGGACACATTTGGTAACTGGCCTT	1623	
2160	CT	TACTGGGAGACATCTGCTTGGGGTTCCTGCTCTGAAATTAATCTTGTGACTGCTCTC	2219	
1624	AAT	GCAACCAATTTGGCTGGAAATGGAGAAACCACTGTGGTGTAGGCATTCAGATCTGGAGA	1683	
2220	AAT	GTGACCAATTTGGCTGGAAATGGAGAAACCACTGTGGGTGGGCATTCAGACACAAAG	2279	
1684	GT	CTTCTGTCTCAAGAGTCACTGGGACAGATTAATGACCAAAAGATGTCAGATTTCTACT	1743	
2280	GT	CTTCTGTATACAGAGTCAATGTGGGCAAGTAATGACCAAAAGATGTCAGAGTCTACA	2339	

1744	QY	CGACCTGGAACCTGTGGCCGCTGTTTCTCCATGCAAAAAGAACTCTATTGTGACTGCT	1803
2340	DB	CGCGCAGAAACTGTAAAGACCTGTGTTTCTCCATGCAAAAGAACTGCTGTGACTGCT	2399
1804	QY	TTCACTGAGTGGACACCTGTGCCAAGGATGTGCCAAGCAGGAATGCCAGTAAACAG	1863
2400	DB	TTCACTGAGTGGACACCACTGCCCGGCTGCCAGCCAGGAAATATCTACAATAAAACAA	2459
1864	QY	TTCTCGATACAGAACTCATCTCCAAAGACGCCAATGGAGGCCAGGAATGCCAGATACC	1923
2460	DB	TTCTCGATACAGGAATTTATCATTCAGGAACAGCAATGGGGTCAAGATGCGCCAGACAC	2519
1924	QY	TTTATATGAGGACAGAGAGTGTGAAGATGTTTCTTTGTCTCTGTATATTCGTTGGAAGCCA	1983
2520	DB	TTGTTTGAAGAAAGAGAGTGTGAAGATATCTCACTGTGTCTCTTATACCGGTGGAAGCCG	2579
1984	QY	CAGAAATGAGCCCTTGCACTTTAGTGCAGAGTCTGTCTGGCAGGGAATAACGGGCGAGC	2043
2580	DB	CAGAAATGAGAGTCTTTGTATCTGTTTCCAGATCTATCAGGCAGGGGAGAACAGGCACC	2639
2044	QY	AGTGAAGCCTGTGGAAAGGGGTTACAAACAAGAGTGTCTCATGATCTCTGTATGACAAAC	2103
2640	DB	AGTGAGGCATGTGGGAAGGGCTGCAGACAAGAGCTGTCTCATGCTCTCTGATGATTAAC	2699
2104	QY	CGGTCAAGCAAAATGATGAAATGCGCTCTAGCAGACACAAACGGCATCCCTCTCTCTGTGCAA	2163
2700	DB	CAGTCTGCGCAAAATGACAGATGCGCTCTAAGCAGATGMAATGGCATGCTTCCCTTGTGCGAG	2759
2164	QY	GAATGCAACAGTCCCATGTGTGGAAGAGCTGCACCTTCACTGCTTGTGTCAAGTTTACGGCC	2223
2760	DB	GAATGCAACCATTTCCCTGTGTGGAGCAGACTGCACCTTCAACCCCTTGGTCTAAAGTTCACTCC	2819
2224	QY	TGCTCCAGCAAGTGTGAAGCCACAAAGATAGCGCGCAGACACTCACAGGGGAAAGACAGA	2283
2820	DB	TGCTCCAAGAACTGTGAAGCAACTCAAGATCAGAGCGCGCAGCTTAACAGGGAAGACAGG	2879
2284	QY	AAGAAGGAGAAATGCCAGGATTTGACCTTTACCTCTAGTGGAGACAGAACTATGTCTCT	2343
2880	DB	AAGAAAGAGAAGTGCCAGGATGCGTCCCTGTACCCCTCTGGTGGAAAATGGAACCATGTCCC	2939
2344	QY	TGTGATGAATTTATATCCCAACCTTATGGAACCTGTGCAAGATTTGCAATCTTCCAGAAAGCC	2403
2940	DB	TGTGATACGTTTATGTCTCATCTCTTATGGAACCTGGTCACTTTCATTCACAGAGGC	2999
2404	QY	AGAAGGAGCGCTCAACC---GAGGACTCGGGTACAAGCAGACAGCAAGAAAGATGTGAGAA	2460
3000	DB	AAAAGGAGTGTCTCAGCAAGGAGGATTTGGGTGCAAGGAGATGACAAAGAAATGTGAGAA	3059
2461	QY	GGCCTGGCTTTTCAGACAGTAGCCTGTTCTGATAAAAATGGAAGACCTGTGTGACCCCTCC	2520
3060	DB	GGCCTGGCTTTTCAGCAATAGCATGCTCCAAATATAATGGAAGACCGGTTGACCCCTCA	3119
2521	QY	TTCTGCGACAGCTCTGCTTTACATTTCAAGAAAAATGTGTCTATTCCTCGCCATTTGATTCG	2580
3120	DB	TTCTGCGACAGCTCTGCTTTACATTTCAGAAAGCATGTGTCTATCCCTGTCCGTTTGTATTCG	3179
2581	QY	AAGTTAAGCGAATGGTCTAGTTGGGGTCTTGCAGTTTCATCTTGTGAAATGTGAGATGAGA	2640
3180	DB	AAGCTTAAGTGACTGGTTCAGCTGGGGTCTTGTGAGCTCATCTTGTGGGATTTGAGTGAAGA	3239
2641	QY	ATTTCGATCCAAATGGCTAAAAGAAAAACCTTTACAATGGAGGAGCAGCATGTCCCAAACTG	2700
3240	DB	ATTTCGATCCAAATGGCTAAAAGAAAAACCTTTACAATGGAGGAGCAGCATGTCCCAAACTG	3299
2701	QY	GATCTCAAGAAATCAGGT	2717
3300	DB	GATCTCAAGAAATCAGTT	3316

RESULT 2	AK083798	3242 bp	linear	HTC 20-SEP-2003
LOCUS	AK083798		mRNA	



DEFINITION Mus musculus 12 days embryo spinal ganglion cDNA, RIKEN full-length enriched library, clone: D130011L22 Product: weakly similar to HYPOTHETICAL PROTEIN D1110N13.1 IN CHROMOSOME 7 (FRAGMENT) [Homo sapiens], full insert sequence.

ACCESSION AK083798

VERSION AK083798.1 GI:26101501

KEYWORDS HTC; CAP trapper.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS Carninci, P. and Hayashizaki, Y.

TITLE High-efficiency full-length cDNA cloning

JOURNAL Meth. Enzymol. 303, 19-44 (1999)

MEDLINE 99279253

PUBMED 10349636

REFERENCE 2

AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL Genome Res. 10 (10), 1617-1630 (2000)

MEDLINE 20499374

PUBMED 11042159

REFERENCE 3

AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsu, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, K., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer

JOURNAL Genome Res. 10 (11), 1757-1771 (2000)

MEDLINE 20530913

PUBMED 11076861

REFERENCE 4

AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.

TITLE Functional annotation of a full-length mouse cDNA collection

JOURNAL Nature 409, 685-690 (2001)

REFERENCE 5

AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL Nature 420, 563-573 (2002)

REFERENCE 6

AUTHORS Nature 420, 563-573 (2002)

TITLE 6 (bases 1 to 3242)

JOURNAL Nature 420, 563-573 (2002)

AUTHORS Adachi, J., Aizawa, K., Akimura, T., Atakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toyota, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

TITLE Direct Submission

JOURNAL Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN.

Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.go.jp/  
URL: http://fantom.gsc.riken.go.jp/.

FEATURES

Location/Qualifiers

1..3242

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="FANTOM\_DB:D130011L22"

/db\_xref="MGI:2419369"

/db\_xref="taxon:10090"

/clone="D130011L22"

/tissue\_type="spinal ganglion"

/clone\_lib="RIKEN full-length enriched mouse cDNA library"

/dev\_stage="12 days embryo"

misc\_feature

1..3242

note="weakly similar to HYPOTHETICAL PROTEIN D1110N13.1 IN CHROMOSOME 7 (FRAGMENT) [Homo sapiens] (SPTRIO43376, evidence: FASTY, 53.6%ID, 99.4%length, match=540)"

ORIGIN

Query Match 43.4%; Score 1908.8; DB 11; Length 3242;

Best Local Similarity 82.7%; Pred. No. 0;

Matches 2196; Conservative 0; Mismatches 457; Indels 3; Gaps 1;

4 GTGCGCTGCATTTCAGAACTGAAACGAACTGTGGTGCAGAAATGAAATGCGAACACTTT 63

586 GTGCGCTGTCTTCAGAACTCAATAGAACCACTGTGTCCATGATGATCTGTGAACACTTT 645

64 GCCTTCAGCTCCCTACAGACAGCGCTTGCCTTCATCTCTCCCGGATTTGTGTATA 123

646 GTCTCTCAGCCCCCAGAGCAAGCTTCCCTCATCTCTCCCGGATTTGTGTATA 705

124 TCTGAGTTTCTTACCATGTGTCCAACTGTAGCAAGGATGTGGAGAAATTCAGCATAGA 183

706 TCTGAGTTTCTCCCATGTGTCCAACTGTCCCGGAGGATGTGGAGAAATTCAGCATAGA 765

184 ACTCGGGGTATAGCTGCGCCCTCTCTTTGGTGGTTCGAATGTCCTCAATCTGACTGAG 243

766 ACTCGTGTGGCCATCGCGCCCTCTCTGTATGAGGCTGTGCAGTGTCCAAATCTCAGGAG 825

244 TCAAGAGCTGTGATGCTCCCAATTTCTCTCTCTCTGCGGAGAGAGAAATATATCAATTAGC 303

826 TCCAGAGCTGTGAGGCTCCAGTTTCTCTCTCTCTGCGGAGAGAGAAATATATCTCTCAGC 885

304 CTTAAGTTGGACCATGGAGTAAATGCGAGCTGCTCTCTCTTAAAGAAATTAATTCAGAC 363

886 CTAAGAGTGGGACCATGGAGCAATGTAGACTCCCTCACTTAAGGAAAGTGCACCTCAGC 945

364 GGAAGAACTGTTCTGGATTTTAACTCTGANTTCAATGAGCGAGTCACTTTTAAACATCAA 423

946 GGAAGAAATATCCAGATTTTGTAGTCTGANTTCAACAGAGCAAGTCACTTCAACATCAG 1005

424 AGTTACAAAGCAATCATCATTCGAGGCTTGGGCAATAGAGATAGGTTATCAAAACCCGG 483

1006 AGTTACAAAGCAATCACTCCAGCTTCCAGGCTGTGGGATGTAGTATAGGTTTCAAAACCCAG 1065

484 CAGGTTTCTGTACAGAGTGCATGCAAAATGCTATGTTTAAAGCTTTGCTTCAAGAT 543

1066 CAGGTTTCTGTACAGAGTGCATGCAAAATGCTATGTTTAAAGCTTTGCTTCAAGAT 1125

544 TCCTTCCCAATGACTGTTTCAAGTCTTCATGCTCCCAAGACTGTGAACCTCCAGTGG 603

1126 TCCTTCCCTCTGACTGTCCAGCTTGTGTGATGCCAAAGACTGTGAGACTTCTCAATGG 1185

604 TCCTTCTGAGCCCTGCTCCAGACATGCTCCCTTCCAGGAGTCTCTTCCAGGATTTAGG 663

1186 TCCTTCTGAGCCCTGCTTCCAGACCTGCTCCCTTCTGGGACTTTTACCTCGAGTTAGA 1245

664 AGCAGAGCGGAACTGTGAAGCAGCATGCTATTTGGAGTGGAAAGAGTGTCTCAACTT 723

1246 AGCAGGAGCCGGAACGTGAAGCATATTGCTATTGGAGGTGGCCAGGAATGCCAGAACTT 1305  
724 CTTGAGAAAGAGGCGCTCATTTGTTGAAGGAGAACTTCTGCAGCAATGTGCCAGGTATTC 783  
1306 CTTGAGAAAGAGACCTGCTATGCTGAAGGGAATTTCTGCAACCATGTCCAGGTATTC 1365  
784 TGGAGAACTTCTGAATGGAAGAAATGCAAGTCTCTCTCTCTCTCGAGCAGCAGGATCCC 843  
1366 TGGAGAACTTCTGAATGGAAGAAATGCAAGTCTCTCTCTCTCTCGAGCAGCAGGATCCC 1425  
844 CACTGGCATGTGACGGGACCGTGTGGCGGTGGGATCCAGACCCGGGAGGTGTACTGT 903  
1426 CTTTGGCATGAGACAGACCAATCTGTGTGTGTGGATCCAGACCCGGGAGGTGTACTGT 1485  
904 GCCCAGAGCTACAGCAGTGGCGGACTGAGGGCCAAAGAACTCTTAGACCTGTGGAA 963  
1486 GCCCAGAGCTACAGCAGTGGCGGACTGAGGGCCAAAGAACTCTTAGACCTGTGGAA 1545  
964 AAGGCATTATGTGGGACCGCCCGTGTGCGCTCTCAGCTCTGCAATATCCCTGCTCT 1023  
1546 AGACACTATGCTTGGGACCTGCCCCCTCAGCCTCTCAGCTTGTGCAATGTCCCTTCTCG 1605  
1024 ACGGACTGCTATGATCTTCTCTGCTCAGCCTGGGCGCTGTGCATCCATGAAAATCTCAT 1083  
1606 ATGACTGCTATGATCTTCTCTGCTCAACTGGGCCCATGTGTGTTTGGAACTGCCAT 1665  
1084 GAACCTCAGGGGAAAAGAAATTTAGAACAGCAGCGGCCCATGCTCTCATGGAATCTACA 1143  
1666 GATCCTCAGGGGAAAAGAACTTTAGAAATGAGACAGGGCCATGATCTCATGGAATCTACA 1725  
1144 GGGCCTGACGGGCAATTTGCCCTCATTTGGTGGAGTCTGTTCTGTGAGGATCCAAATGTGC 1203  
1726 GGGCCCATGGGCGCTGCCCTCATCTGGCAGATCTGTACCTTGTGAGGATCCAAATGTGT 1785  
1204 TACCGATGGCTGGCATCAGAAAGGAACTGTTTCCCTGATCATGGAATAATGTGCCCTGGGA 1263  
1786 CATCGCTGGCTTGGCTCGGAAGGAACTGTATTGCTCATCATGGGAAATGTGCCCTGGGA 1845  
1264 CATCGTATTCTGAAGGCGCTCCAGAAATGACCGGGAGAGATGATATCAGGGAGTCTT 1323  
1846 CATCGCATCTGAAGAGCTGTCTGTCAGAAATGACCGGGAGAGAAATATCAGGGGCTCTC 1905  
1324 TGCCAGTTCCT 1383  
1906 TGTCCAGTGTCT 1965  
1384 GTGCTGAGCGAGTGGACGAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1443  
1966 GTAGTGTGAGTGGACGAGTGTCT 2025  
1444 GATGGGAAACAGACCCAGGTCAAGAACTATCTCTGGCCTCTGGCTGGGAAAGTGGAAAGCCA 1503  
2026 GATGGGAAACAAACCCAGGTCAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2085  
1504 TGTCT 1563  
2086 TGGCT 2145  
1564 CTTCACTGGGAGCATCGCTTGGGCGCTTGTCTGAGGACATTTGTTAATGCGCTT 1623  
2146 CTCTACTGGGAGCATCTGCTTGGGCTTCTGCTCTGAAATACTCTTGTGACTGTCTCTC 2205  
1624 AATGCAACCATTTGGCTGGAATGAGAAAGCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1683  
2206 AATGTGACCATTTGGCTGGAATGAGAAAGCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2265  
1684 GTCT 1743  
2266 GTCT 2325  
1744 CGACCTGAAACTGTGGCCCT 1803  
2326 CGGCCGAAACTGTGAAGACCTTGTCT 2385

QY 1804 TTCACTGAGTGGACACACCTGTCCCAAGGATGTCCAAGCAGGAATGCCAGTAAACAG 1863  
Db 2386 TTCACTGAGTGGACACACCTGTCCCAAGGATGTCCAAGCAGGAATGTACTATTAACAA 2445  
QY 1864 TCTCGATACAGAACTCATCATCAAGAGAGCAGCCAAATGAGGCGCAGGAATGCCAGATACC 1923  
Db 2446 TCTCGATACAGAACTCATCATCAAGAGAGCAGCCAAATGAGGCGTCAAGAGTGCAGACACC 2505  
QY 1924 TTATATAGGAGAGAGAGTGTGAAGATGTTTCTTGTGTCTGTATATCGGTGGAAGCCA 1983  
Db 2506 TTGTTTGAAGAAAGAGAGTGTGAAGATATCTCACTGTGTCTTCTATCCGTGGAAGCCG 2565  
QY 1984 CAGAAATGGAAGCCCTTGCATCTTATAGTCCAGAGTGTGTGTGGCAGGAATAACCGGCGAGC 2043  
Db 2566 CAGAAATGGAAGCCCTTGTATCTTGTATCCAGTGTCTATCAGGACAGGGGAGAAACAGGCGACC 2625  
QY 2044 AGTGAAGCCTCTGGAAGGGGTATCAACAGAGAGTGTCTCATGCTCTCTCATGCTCTCTGATGACAA 2103  
Db 2626 AGTGAAGCCTCTGGAAGGGGTCTGAGCAAGAGCTGTCTCATGCTCTCTCTGATGATTAAC 2685  
QY 2104 CGGTGACAGAGAAATGATGGAATGCTCAAGCAGACAAACCGCATGCTCTCTCTTGTGCAA 2163  
Db 2686 CAGTCTGACAGAAATGACAGAGTGTCTCAAGCAGATGAATGGCATGCTCTCTCTTGTGCG 2745  
QY 2164 GAATGACAGTCCCATGTCGAGAGAGTGTGACCTTCTCACTGCTTGTGCTCAAGTGTACGCCC 2223  
Db 2746 GAATGACACCATTTCCCTCTGCGGACGACTGCACTTCACTCTCTCTCTCTCTCTCTCTCTCT 2805  
QY 2224 TGCTCCACGAACTGTGAAGCCACAAAAGTAGGGGCGGACAGCTCACAGGGAAGAGAGA 2283  
Db 2806 TGCTCCACGAACTGTGAAGCAACTCAGATCAGGAGGCGGAGCTAAACAGGGAAGAGAGG 2865  
QY 2284 AAGAGGAGAAATGCCAGGATTTGACCTTTTACCTCTAGTGGAGACAGAACTATGTCTCT 2343  
Db 2866 AAGAGGAGAAATGCCAGGATTTGACCTTTGACCTCTGCTGGAATGGAACCTATGTCTCT 2925  
QY 2344 TGTGATGAATTTATATCCACCTTATGGAACCTGTCAGATTCATTTCTTCCAGAGGC 2403  
Db 2926 TGTGATGATTTATGTCATCTCTTATGGAACCTGTCAGCTTGCATTTCTACCAAGAGGC 2985  
QY 2404 AAGAGGAGCCTCACCC--GAGGACTCGGGTCAAGCAGACAGCAAGAAATGTGGAGAA 2460  
Db 2986 AAGAGGAGCTCAGCAAGAGAGGATTTGGGTGCAAGAGATGACAAAGAAATGTGGAGAA 3045  
QY 2461 GGCTCTGGCTTTCCAGCAGTAGCTGTCTGATAAATAATGGAAGACCTGTGAGCCCTCC 2520  
Db 3046 GGCTGCGCTTTCCAGCAATAGCATGCTTCAATATAAATGGAAGACCGGTGTGAGCCCTCA 3105  
QY 2521 TTCTGCGAGCAGCTCTGGTTTACATTTCAAGAAAAATGTGTCTATTTCCCTGCCCATTTGATTGC 2580  
Db 3106 TTCTGCAACAGCTCTGGTTTACATTTCAGGAAGCATGTGTCTATCCCTGTCTCTGATTGC 3165  
QY 2581 AAGTTAAGCAGATTGGTCTAGTTGGGGGTCTTGCAGTTCTCTTGTGGAATTTGAGTGAGA 2640  
Db 3166 AAGTTAAGTGTAGTGGTCCAGCTGGGGTCTTGTGAGCTCATCTTGTGGGATTGGAGTGAGA 3225  
QY 2641 ATTCGATCCAAATGGC 2656  
Db 3226 ATTCGATCCAAATGGC 3241

## RESULT 3

AK047508

LOCUS

DEFINITION

AK047508 2266 bp mRNA linear HTC 20-SEP-2003  
Mus musculus 10 days neonate cerebellum cDNA, RIKEN full-length  
enriched library, clone:B930082a18 product:weakly similar to  
HYPOTHETICAL PROTEIN D1110N13.1 IN CHROMOSOME 7 (FRAGMENT) [Homo  
sapiens], full insert sequence.

ACCESSION

AK047508

VERSION

AK047508.1

KEYWORDS

HTC; CAP trapper.

SOURCE

Mus musculus (house mouse)

ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1  
 Carninci, P. and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning  
 Meth. Enzymol. 303, 19-44 (1999)  
 99279253 PUBLISHED  
 10349636  
 2  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
 Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes  
 Genome Res. 10 (10), 1617-1630 (2000)  
 20499374 PUBLISHED  
 11042159  
 3  
 Shibata, K., Itoh, M., Aizawa, K., Nagao, K., Sasaki, N., Carninci, P.,  
 Kono, H., Akiyama, J., Nishi, K., Kitsu, T., Tashiro, H., Itoh, M.,  
 Sun, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,  
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J.,  
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer  
 Genome Res. 10 (11), 1757-1771 (2000)  
 20530913 PUBLISHED  
 11076861  
 4  
 The RIKEN Genome Exploration Research Group Phase II Team and the  
 PANTOM Consortium.  
 Functional annotation of a full-length mouse cDNA collection  
 Nature 409, 695-690 (2001)  
 5  
 The FANTOM Consortium and the RIKEN Genome Exploration Research  
 Group Phase I & II Team.  
 Analysis of the mouse transcriptome based on functional annotation  
 of 60,770 full-length cDNAs  
 Nature 420, 563-573 (2002)  
 6 (bases 1 to 2266)  
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M.,  
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hirao, T., Hirozane, T.,  
 Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kacawa, I., Kasukawa, T.,  
 Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,  
 Koyu, S., Kurinara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
 Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,  
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,  
 Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T.,  
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, P., Takaku-Akahira, S.,  
 Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
 Muramatsu, M. and Hayashizaki, Y.  
 Direct Submission  
 Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of  
 Physical and Chemical Research (RIKEN), Laboratory for Genome  
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
 Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,  
 URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,  
 Fax: 81-45-503-9216)  
 cDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues.  
 Please visit our web site for further details.  
 URL: http://genome.gsc.riken.go.jp/  
 Location/Qualifiers  
 1. .2266  
 /organism="Mus musculus"

misc\_feature 1. .2266  
 /note="weakly similar to HYPOTHETICAL PROTEIN DJ1110N13.1  
 IN CHROMOSOME 7 (FRAGMENT) (Homo sapiens) (SPTRO43376,  
 evidence: FASTV, 53.8%ID, 99.4%length, match=540)"  
 ORIGIN  
 Query Match 15.2%; Score 666.4; DB 11; Length 2266;  
 Best Local Similarity 81.4%; Pred. No. 7.5e-195;  
 Matches 772; Conservative 0; Mismatches 176; Indels 0; Gaps 0;  
 QY 4 GTGCGTGCATTCAGAGCTGACGAACTGGTGTGCAATGAATGAATGCGACACTTT 63  
 DB 544 GTGCGTGTCTTTCAGAGCTCAATGAGACCACTGGTGTCAATGAGATCTGTGAACACTTT 603  
 QY 64 GCGCTTCAGCCTCTACAGAACAGGCTTGCCTCATTCCTTGTCCCGGATTTGTAGTA 123  
 DB 604 GCTCCTCAGCCCCCAGAGCAAGCTGCTCTCATTCCTTGTCCCGGATTTGTAGTA 663  
 QY 124 TGTGAGTTCTTACCATGTCCTCAACTGTAGCAAGGATGTGGAGAAATTCGACGATAGA 183  
 DB 664 TGTGAGTTCTTCCCATGTTCCACCTGCCCCAGGATGTGGAGAAATTCGACGATAGA 723  
 QY 184 ACTGCGCGGTGCATAGTCTCCCTCTCTTTGTGTGTTTGCATGTCTCAAAATCTGACTGAG 243  
 DB 724 ACTGCTGTGGCATCGGCCCTCTGTATGAGGCTGTGAGTGTCTCAAAATCTCAGGAG 783  
 QY 244 TCAAGAGCTGTGATGCTCCCATTTCTCTCTCTTGGGAAAGAGAAATATACATTTAGC 303  
 DB 784 TCAGAGCTGTGAGGCTCCAGTTTCTCTCTCTTGGGAAAGAGAAATATATTTCTTCTCAGC 843  
 QY 304 CTTTAAAGTTGGACCTGGAGTAAATGACAGACTGCTCATCTTTAAAGAAATTAATCAAGC 363  
 DB 844 CTAAAGTTGGACCTGGAGCAATGTAGACTCTCTCACTTAAAGAGTCCACCTCAGC 903  
 QY 364 GGAAGAACTGTTCTGAGTTTAACTCTGATTCATTAAGAGCGAGTCACTTTAAACATCA 423  
 DB 904 GGAAGAAATATCAGGATTTAGCTCTGATTCATTAAGAGCAAGTCACTTCAACATCAG 963  
 QY 424 AGTTACAAAGCAATCATCATTCGAGTCTTGGCAATAGAGATAGGTTTATCAAAACCGG 483  
 DB 964 AGTTACAAAGCAATCATCATTCGAGTCTTGGCAATAGAGATAGGTTTATCAAAACCGG 1023  
 QY 484 CAGGTTTCGTGTACAAAGAGTGTATGACAAATGCTATGTTAAGCCTTTGCTTCAAGAT 543  
 DB 1024 CAGGTTTCGTGTACAAAGAGTGTATGACAAATGCTATGTTAAGCCTTTGCTTCAAGAT 1083  
 QY 544 TCCTTCCCATGACTGTTTCAGTCTGTCATGTCGCAAGAGTGTGAACCTCCAGTGG 603  
 DB 1084 TCCTTCCCTCTGACTGTCCAGCTTGTGTGATGCCAAAGAGTGTGAGACTTCTGAATGG 1143  
 QY 604 TCCTTCTGAGGCCCCCTGCTCCAGAGATCCCGTTCAGGAGTCTCTTGGCAGGATTTAGG 663  
 DB 1144 TCCTTCTGAGGCCCCCTGCTCCAGAGATCCCGTTCAGGAGTCTCTTGGCAGGATTTAGG 1203  
 QY 664 AGCAGAGCGGAACTGTAAGCAATGCTGATTTGGAGTGGAAAGAGTGTCTGACTTT 723  
 DB 1204 AGCAGAGCGGAACTGTAAGCAATGCTGATTTGGAGTGGAAAGAGTGTCTGACTTT 1263  
 QY 724 CTTTGAGAAAGAGGCTGCTGATTTGTTGAAGGAGAACTTCTGACGAATGTCCAGGATTTCC 783  
 DB 1264 CTTTGAGAAAGAGGCTGCTGATTTGTTGAAGGAGAACTTCTGACGAATGTCCAGGATTTCC 1323  
 QY 784 TGGAGAACTTCTGAATGGAAGAAATGCCAAGTCTCTCTCTCTCTCGAGGAGGATCCCC 843

FEATURES  
 source



Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN		Query Match	13.7%	Score	600.4	DB	10	Length	760
		Best Local Similarity	95.7%	Pred.	No. 1.3e-174				
		Matches	639	Conservative	0	Mismatches	26	Indels	3
		Gaps	2						
QY	354	TAATCCAGCGGAGAACTGTTCTGGATTTTAACTCTGATTCATCAATGAGCGAGTCACCTT	413						
Db	1	TAATCCAGCGGAGAACTGTTCTGGATTTTAACTCTGATTCATCAATGAGCGAGTCACCTT	60						
QY	414	TAAACATCAAGTTACAAAGCACATCATTCGAAAGTCTTGGCAATAGAGATAGTTA	473						
Db	61	TAAACATCAAGTTACAAAGCACATCATTCGAAAGTCTTGGCAATAGAGATAGTTA	120						
QY	474	TCAACCCGGCAGGTTTCGTGTACAAAGAGTGAGCAAAATCTATGTTAAGCCCTTTG	533						
Db	121	TCAACCCGGCAGGTTTCGTGTACAAAGAGTGAGCAAAATCTATGTTAAGCCCTTTG	180						
QY	534	CTTCAAGATTCCTTCCCATTTGCTGATTCAGTCTGATCATGCCCCAAAGACTGTGAAC	593						
Db	181	CTTCAAGATTCCTTCCCATTTGCTGATTCAGTCTGATCATGCCCCAAAGACTGTGAAC	240						
QY	594	CTCCAGTGTCTCTCGAGGCCCTGCTCCAGACATGCCGTTTCAAGGAGTCTCTTGCC	653						
Db	241	CTCCAGTGTCTCTCGAGGCCCTGCTCCAGACATGCCGTTTCAAGGAGTCTCTTGCC	300						
QY	654	AGGATTTAGGAGCAGGAGCCGGAAGCTGAAGCACATGCTATTGGAGTGGAAAGGAGTG	713						
Db	301	AGGATTTAGGAGCAGGAGCCGGAAGCTGAAGCACATGCTATTGGAGTGGAAAGGAGTG	360						
QY	714	TCTGAACTTCTTGAGAAAGAGCCCTGATTTGAGAGGAACTCTGAGCAATGTCC	773						
Db	361	TCTGAACTTCTTGAGAAAGAGCCCTGATTTGAGAGGAACTCTGAGCAATGTCC	420						
QY	774	CAGGTATTCCTGGAGACTTCTGAATGGAAGATGCCAAGTCTCTCTCTCTCGAGCA	833						
Db	421	CAGGTATTCCTGGAGACTTCTGAATGGAAGATGCCAAGTCTCTCTCTCTCGAGCA	480						
QY	834	GCAGGATCCCCACTGGCATGTGACGGGACCCCGTGTGTGGCGGTGGATCCAGACCCGGG	893						
Db	481	GCAGGATCCCCACTGGCATGTGACGGGACCCCGTGTGTGGCGGTGGATCCAGACCCGGG	539						
QY	894	GTGTACTGTGCCCCAGAGCTACAGCAGCTCCCGACTGAGGCGACGAGTCTCTAG	953						
Db	540	GTGTACTGTGCCCCAGAGCTACAGCAGCTCCCGACTGAGGCGACGAGTCTCTAG	599						
QY	954	ACCTGTGGAAGGCAATATGTGTGGGACCCCGCCCGTGTGCTCTCAGCTCTGCAATAT	1013						
Db	600	ACCTGTGGAAGGCAATATGTGTGGGACCCCGCCCGTGTGCTCTCAGCTCTGCAATAT	657						
QY	1014	CCCTTGCT 1021							
Db	658	CCTGGCT 665							

RESULT 6  
CF729699  
LOCUS  
DEFINITION  
UI-M-HO-clh-1-13-0-UI.r1 NIH BMAP\_HEO Mus musculus cDNA clone  
IMAGE:30631404 5', mRNA sequence.  
CF729699  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE

CF729699 772 bp mRNA linear EST 09-OCT-2003  
UI-M-HO-clh-1-13-0-UI.r1 NIH BMAP\_HEO Mus musculus cDNA clone  
IMAGE:30631404 5', mRNA sequence.  
CF729699  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE

CF729699.1 GI:37603867  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 772)  
NIH-MGC <http://mgc.nci.nih.gov/>  
National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL COMMENT

Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue procurement: Dr. James Lin Unive  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Distribution information can be found at  
<http://genome.uiowa.edu/distribution/mousefi.html>  
This clone was contributed by the Brain Molecular Anatomy Project  
(BMAP)

Seq primer: pYX-5

Location/Qualifiers

FEATURES

source

1. .772  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="CS7BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30631404"  
/tissue\_type="whole eye"  
/dev\_stage="embryo 12.5,13.5,14.5 dpc"  
/lab\_host="DH10B (T1 phage resistant)"  
/clone\_lib="NIH\_BMAP\_HEO"  
/note="Organ: Eye; Vector: pYX-Asc; Site 1: Ecor I;  
Site 2: Not I; The library was constructed according  
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
1996. Denatured RNA was size fractionated on a 1% agarose  
gel. First strand cDNA synthesis was primed with oligo-dT  
primer containing a Not I site. Double strand cDNA was  
size selected according to mRNA size fraction, ligated  
with Ecor I adaptor, digested with NotI and then cloned  
directionally into pYX-Asc vector. The library tag  
sequence located between the Not I site and the polyA tail  
is TTAATGAGT. This library was created for the University  
of Iowa Brain Anatomy Project (BMAP). Gene Discovery in the  
Developing Mouse Nervous System", supported by National  
Institute of Mental Health (NIMH)."

ORIGIN

		Query Match	12.9%	Score	565.2	DB	14	Length	772
		Best Local Similarity	83.2%	Pred.	No. 1.2e-163				
		Matches	642	Conservative	0	Mismatches	130	Indels	0
		Gaps	0						
QY	1232	GTTTCCCTGATCATGAAATATGGCTGGGACATCGTATTCTGAAGCCGCTCCAGCA	1291						
Db	1	GTTTCCCTGATCATGAAATATGGCTGGGACATCGTATTCTGAAGCCGCTCCAGCA	60						
QY	1292	ATGACCGCGAGAGATGATCAGGGAGTCTTGGCCAGTCCCTCTCTCTGAGAGCA	1351						
Db	61	ATGACCGCGAGAGATGATCAGGGAGTCTTGGCCAGTCCCTCTCTCTGAGAGCA	120						
QY	1352	AGTCTTGTGAATTCCTGCGCAATGGACTGTGTCTGAGCGAGTGGACGAGTGTCTAT	1411						
Db	121	TGGCTTGTGAATTCCTGCGCAATGGACTGTGTCTGAGCGAGTGGACGAGTGTCTAT	180						
QY	1412	CTGTGTTCCAGTCTCTTCAAAATAAAACTCAGATGGGAAACAGACCAGGTCAAGACTA	1471						
Db	181	CTGTGTTCCAGTCTCTTCAAAATAAAACTCAGATGGGAAACAGACCAGGTCAAGACTA	240						
QY	1472	TCCTGGCACTGGCTGGGAGGTGGAAGCCATGTCCCTAGTCTAGGCTCTCCAAAGAGC	1531						
Db	241	TCCTGGCACTGGCTGGGAGGTGGAAGCCATGTCCCTAGTCTAGGCTCTCCAAAGAGC	300						
QY	1532	ATCGTTTGTGTAATGACCAATTCCTGTATGACGTTTCACCTGGGAGACATCGCTTGGGGC	1591						
Db	301	ACCCTATGCAATGATCATTCCTGTACACAGCTCTACTGGGAGACATCTGCTTGGGTT	360						
QY	1592	CTGTGTTCTGAGGACATGCTGTAATCCCTTAATGCAACCATTTGGCTGGATGGAGAAG	1651						
Db	361	CTGTGTTCTGAGGACATGCTGTAATCCCTTAATGCAACCATTTGGCTGGATGGAGAAG	420						
QY	1652	CCAGTGTGTGTAGGCATTCAGACTCGGAGAGTCTTCTGTGTCAAGAGTCAAGTGGGAC	1711						





REFERENCE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 820)  
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)

Tissue Procurement: Dr. Jim Lin, University of Iowa  
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 This clone was contributed by the Brain Molecular Anatomy Project  
 (BMAP)

Seq primer: pYX-5.

Location/Qualifiers

SOURCE

1..820  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone="IMAG:6405956"  
 /tissue\_type="whole brain"  
 /dev\_stage="embryo 12.5dpc"  
 /lab\_hosts="DH10B (TI phage resistant)"  
 /clone\_lib="NIH BMAP\_F00"  
 /note="Organ: Brain; Vector: pYX-Asc; Site: 1: EcoR I;  
 Site 2: Not I; The library was constructed according  
 Bernaldo, Lemmon and Soares, Genome Research, 6:791-806,  
 1996. Denatured RNA was size fractionated on a 1% agarose  
 gel. First strand cDNA synthesis was primed with oligo-dT  
 primer containing a Not I site. Double strand cDNA was  
 size selected according to mRNA size fraction, ligated  
 with EcoR I adaptor, digested with NotI and then cloned  
 directionally into pYX-Asc vector. The library tag  
 sequence located between the Not I site and the polyA tail  
 is TGAGAGACC. This library was created for the University  
 Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the  
 Developing Mouse Nervous System', supported by National  
 Institute of Mental Health (NIMH), Hemlin Chin, Ph.D.,  
 program coordinator."

ORIGIN

Query Match 12.3%; Score 542.6; DB 13; Length 820;  
 Best Local Similarity 81.5%; Pred. No. 1.4e-156;  
 Matches 626; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

QY 395 CAATGACGAGTCACCTTTAAACATCAAGTTACAAAGCACATCATTCATTCGAAGTCCT 454  
 DB 1 CAACGAGCAAGTCACCTTAAACATCAAGTTACAAAGCACACACCTCCAGCCTG 60  
 QY 455 GGGCAATGAGATAGTTATCAACCCGCGAGGTTCTGTGTACAGAGTGATGGACAA 514  
 DB 61 GGGATGTAGTATAGGGTTTCAACACCGAGGAGGTTGTGTGTACAGAGTGATGGAGAA 120  
 QY 515 ATGCTATGTTAAGCCTTTGCGCTTCAAGATTCCTCCCATTCGACTGTTCAGTCTCGATCA 574  
 DB 121 AGCCCTGTTAAGCCTTTGTTGTGTGATTCCTTCCTCTGACTGTCCAGCTTGTGTGA 180  
 QY 575 TGCCCAAGACATGTGAACCTCCAGTGGTCTCTGTGGAGCCCTGCTCCAGACATGCC 634  
 DB 181 TGCCAAAGACATGTGAGACTTCTGAATGGTCTCCATGAGAGCCCTGTTCCAAAGACCTGCC 240  
 QY 635 GTTCAGGAGTCTCTTGCCAGATTTAGAGCAGAGCGCGAAGCTGAAGCACATGGCTA 694  
 DB 241 GTTCTGGGACTTTATCACTGTGAGTTAGACAGAGCGCGAAGCTGAAGCATTTGCTA 300  
 QY 695 TTGGAGGTGGAAGAGGTGTCCTGAACTTCTTGAGAAAGAGCGCTGATGTTGAAGGAG 754  
 DB 301 TTGGAGGTGGCCAGGAATGCCAGAACTTCTTGAGAAAGAGACCTGATGTTGAAGGGG 360

QY 755 AACTTCTGACGATGTCCTCCAGGATATCTCTGGAGAACTTCTGAATGGAAGAAATGCCAAG 814  
 DB 361 AATTCTTGCAACCATGTCCAGGATATCTCTGGAGAACTTCTGAATGGAAGAAATGCCAAG 420  
 QY 815 TCTCTCTCTCTCTCGACGACGAGTATCCCACTGGCATGTGACGGGACCCGCTGTGTGGCG 874  
 DB 421 TCTCTCTCTCTCTCGACGACGAGTATCCCACTGGCATGTGACGGGACCCATCTGTGTGTG 480  
 QY 875 GTGGATCCAGACCCGGGAGGTGTACTGTGCCAGAGCGTACAGAGCTGTGCCCATGA 934  
 DB 481 GTGGATCCAGACCCGGGAGGTGTACTGTGCCAGAGCGTACAGAGCAACCATTCATCAA 540  
 QY 935 GGGCCCAAGAGTCTCTAGACCTCTGGAAAGGAGCATATGTGTGGGACCCGCCCTGTGC 994  
 DB 541 GACCAAGAGTCTCCAGGCGCTGGAAAGGAGCATATGTGTGGGACCTGCCCTCAG 600  
 QY 995 CTTCTAGCTCTGCAATATCCCTTCTCTAGGAGCTGATAGTATCTTCTGTGTCAGCCT 1054  
 DB 601 CTTCTAGCTCTTGAATGTCTCTTCTCGATGAGTATCTTCTGTGTCAACCT 660  
 QY 1055 GGGGCTGTGATCCATCAAAACTGTGATGAACCTCAGGAGGAGGAGGAGGATTTAGAAGA 1114  
 DB 661 GNGGCCCATGTGTTTTTGGAGAACTGCCATGATCTCTCAGGAGGAGGAGGATTTAGAATGA 720  
 QY 1115 GCGAGCGCATGTCTCTCATGAATCTCAGGCGCTCAGGCGCATTCGCC 1162  
 DB 721 GACAGCGCATGTACTCATGAGTCTACAGGCGCATTCAGGCGCGCTGCC 768

RESULT 9

BY735628

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

BY735628 717 bp mRNA linear EST 17-DEC-2002  
 CDNA clone 10C0030112 5', mRNA sequence.

BY735628.1 GI:27148755  
 Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 717)

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,  
 Nikaido, I., Oeato, N., Saito, R., Suzuki, H., Yamanaka, I.,  
 Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,  
 Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,  
 Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,  
 Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V.,  
 Chothia, C., Corbani, D.E., Cousins, S., Della, E., Dragani, T.A.,  
 Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,  
 Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,  
 Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,  
 Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,  
 Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,  
 Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,  
 Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G., Ramchandran, S.,  
 Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramchandran, S.,  
 Ravasi, T., Reed, J.C., Reid, D.J., Reid, J., Ring, B.Z., Ringwald, M.,  
 Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K.,  
 Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,  
 Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,  
 Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,  
 Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,  
 Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,  
 Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,  
 Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,  
 Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,  
 Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,  
 Rogers, J., Birney, E. and Hayashizaki, Y.  
 Analysis of the mouse transcriptome based on functional annotation  
 of 60,770 full-length cDNAs  
 Nature 420, 563-573 (2002)  
 JOURNAL  
 MEDLINE 22354683

TITLE

JOURNAL

MEDLINE

12466851

PUBMED  
COMMENT

12-00051  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222

Sugahara,Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: Real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamakawa,I., and  
Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and  
Hayashizaki,Y.  
Computational Analysis of Full-Length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for  
further details.  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
FEATURES  
source  
1..683  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="D130011122"  
/tissue\_type="spinal ganglion"  
/dev\_stage="12 days embryo"  
/lab\_host="DH103"  
/clone\_lib="RIKEN full-length enriched, 12 days embryo  
spinal ganglion"  
/notes="Site 1: Sali; Site 2: BamHI; cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5'  
GAGAGAGAGCGCGCAACTCGAGTTTCTTTTCTTTTNN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. Second strand cDNA was prepared with the  
primer adapter of sequence [5'  
GAGAGAGAGATTTCGAGTTTAATTAATCCCCCCCCCCC 3']. cDNA  
was cleaved with BamHI and XhoI. Vector: a modified  
pBluescript KS(+) after bulk excision from Lambda FLC I."

361 TGTCCCTGTGATACGTTATGTCTCATCTTATGAACTGGTCAGCTTGCATTCTACCA 420  
2398 GAGGCGAGAGGGAGCTCACC---GAGGACTCGGTACAAGCAGACAGCAAGATGT 2454  
421 GAGGCGAGAGGGAGTGTCTCAGCAAGGAGGATGTGGGTGCAAGGATGACAAAGATGT 480  
2455 GGAGAGGCGCTGCGCTTTTCGAGCAGTAGTACCTGTTCTGATAAAATGGAAGACCTGTTGAC 2514  
481 GGAGAGGCGGTGCGCTTTTCGAGCAATAGCATGTCCAAATATAATGGAAGACCGGTTGAC 540  
2515 CCCTCTTTCGACAGCTCTGTTTACATTCAGAGAAATGTGTCTATTCCTCCCTGCCATTT 2574  
541 CCCCTATTCTTCAACAGCTCTGTTTACATTCAGAGAAATGTGTGTCTCCCTGTCCCTTT 600  
2575 GATTGCAAGTTTAAAGCATTTGTTCTAGTTGGGGGTCTTTGCAGTTCATCTTGTGGAATTTGA 2634  
601 GATTGCAAGCTAAGTGAAGTCTGTCAGCTGGGGTCTTTCGAGCTCATCTTGTGGGATTTGA 660  
2635 GTGAGAAATTCATCAATGTC 2656  
661 GTGAGAAATTCATCAATGTC 682  
RESULT 11  
BI988413 600 bp mRNA linear EST 20-DEC-2001  
LOCUS 3224-89 Mouse E14.5 retina lambda ZAP II Library Mus musculus CDNA,  
DEFINITION mRNA sequence.  
ACCESSION BI988413  
VERSION BI988413.1 GI:17959395  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 600)  
AUTHORS Mu,X., Zhao,S., Pershad,R., Hsieh,T.-F., Scarpa,A., Wang,S.W.,  
White,R.A., Beremand,P.D., Thomas,T.L., Gan,L. and Klein,W.H.  
Gene expression in the developing mouse retina by EST sequencing  
and microarray analysis  
Nucleic Acids Res. 29 (24), 4983-4993 (2001)  
JOURNAL 21671825  
MEDLINE 11812828  
PUBMED  
COMMENT Contact: Klein WH  
Department of Biochemistry and Molecular Biology  
University of Texas M.D. Anderson Cancer Center  
Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA  
Tel: 713 792 3646  
Fax: 713 790 0329.  
FEATURES  
source  
1..600  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/tissue\_type="neural retina"  
/dev\_stage="embryonic day 14.5 post-fertilization"  
/clone\_lib="Mouse E14.5 retina lambda ZAP II Library"

ORIGIN  
Query Match 10.7%; Score 472.6; DB 12; Length 600;  
Best Local Similarity 86.7%; Pred. No. 7.7e-135;  
Matches 520; Conservative 0; Mismatches 80; Indels 0; Gaps 0;  
QY 3065 GGACTTGTGCTGAAGACTCACAGGTGCGAGCTTGCCTCCTCAATGAAATTTGCTTCCAGT 3124  
DB 1 GAATTTGGGGAAGACTCGCAGGTGGCGCCCTGCTCCTCAATGAAATTTGCTTCCAGT 60  
QY 3125 TCCAGTACATCTACAGAGTGGAGCAGCATGCCAGCTGAGTGAACCCACCTCTGTGTC 3184  
DB 61 TCCAGTACATCTGACAGAGTGGAGCAGCATGCCAGCTGAGTGAATTTTCTCTGTGGCC 120  
QY 3185 AAGCGCTCAGGACCCCGCTGCTAAGCTGTGTGAGTGCATGGCAAGCCAGTCAGCATGG 3244

```

121 AGGGTGTAGGACCCGCTGCTGAGCTGTGTGCGAGTGTATGCGAAGTCTGTGACATGG 180
3245 ACCAATGTGAGCAGCATAATTTGGAGAACCCACAGAGATGAGCATTCCTGCTGGTGG 3304
181 ACCACTGCGAGCAGCGTAATCTGGAAGAACCCACAGAGATGAGTATTCCTGCTGGTGG 240
3305 AATGCGTGTCAACTGTGACGCTCTCAGGTTGAGCGCTTGGACAGAGTGTTCACAGACCT 3364
241 AATGTTGTGTCAACTGTGACGCTCTCAGGTTGAGCGCTTGGACAGAGTGTTCACAAACCT 300
3365 GTGGCCATGGAGTCAATGAGCGGACTCGATTTATCATTTATGCAACCCACAGAGAG 3424
301 GTGGCCAAAGAGTGCATGAGCGGACTCGATTTATCATTTATGCAACCCACAGAGAG 360
3425 GAGGCGCATGCCCCACAGAGCTTTACCCAGAGAGAAACCTGCCAGTGACCCCTGCTACA 3484
361 GTGACAAATGCCCCACAGAACTCACCCAGAGAAACCTGCCAGTGACACCTGCTATA 420
3485 GTGGGTCTCTGGCACTGCTGTCATGTAAATGAGAGGTGGAGACTGTGGGAGAGAG 3544
421 GTTGGGTCTCTGGCACTGCTGTCATGTAAATGAGAGGTGGAGACTGTGGGAGAGAG 480
3545 TTCCAGATCCGACCTTTCTGTCATGTCACAGTGTGTTCAATATCTCATGACGCTGGAC 3604
481 TCAGTGTGCGACCTTCTCTGTCATGTCACAGTGTGTTCAATATCTCATGACGCTGGAC 540
3605 GTGTCGAGGATGACCTGTGTGGAGAAATGCCCTTTTCAGGACAGCATCCTGGAAGCAGCTGT 3664
541 CTGTGGAGGAGGCGCTGTGTGGAGAGTACCTTTCCAGAGAAATCCTGGAAGCAGCTGT 600

RESULT 12
AV590588 565 bp mRNA linear EST 27-NOV-2001
LOCUS AV590588 Bos taurus brain fetus Bos taurus cDNA clone E1BR013E03
5', mRNA sequence.
ACCESSION AV590588
VERSION AV590588.1 GI:9701581
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 565)
Takasuga,A., Hirotsune,S., Itoh,R., Jitohzono,A., Suzuki,H., Aso,H.
and Sugimoto,Y.
Establishment of a high throughput EST sequencing system using
poly(A) tail-removed cDNA libraries and determination of 36,000
bovine ESTs
Nucleic Acids Res. 29 (22), E108 (2001)
JOURNAL MEDLINE
PUBMED 21570554
COMMENT 11713328
Contact: Yoshikazu Sugimoto
Animal Genetics Division
Shirakawa Institute of Animal Genetics
Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
Tel: 81-248-25-5641
Fax: 81-248-25-5725
Email: kazusugi@cocococoon.ne.jp
Single pass sequencing.
This clone was obtained from a polyA-deleted cDNA library.
FEATURES
source
1..565
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/clone="E1BR013E03"
/tissue_type="brain"
/dev_stage="fetus"
/lab_host="DH10B"
/clone_lib="Bos taurus brain fetus"

```

```

/notes="Vector: pZL1; Site 1: SalI; Site 2: NotI; Poly A
was deleted from a NotI site"
ORIGIN
Query Match 10.3%; Score 451.6; DB 9; Length 565;
Best Local Similarity 89.3%; Pred. No. 2.6e-128;
Matches 500; Conservative 0; Mismatches 54; Indels 6; Gaps 1;
QY 2259 GCACAGCCTCACAGGAAAAAGCAGAGAAATGCCAGATTCGACCTTTACCC 2318
DB 5 GCGTCCGCTCACAGGAAAAAGGAGAAATGCCAGATTCGACCTTTACCC 64
2319 TCTAGTGGACAGCAACTATGCTTGTGATGATTTATATCCCAACCTTATGGAACTG 2378
DB 65 TCTAGTGGAAACAGAACCATGTCTTGTGATGTTGTATCATCTCATCTTATGAAACTG 124
2379 GTCAAGTTGATCTTCCAGAAAGCAGAGAGGAGCTCACCGAGGACTGGGGTACAAAGC 2438
DB 125 GTCAAGTTGATCTTCTCTGAGAGCAGAAAGAGGCTCATAGAGGACTGCGGCTACAAGG 184
2439 AGACAGCAAAAGATGTGGAGAGGCGCTTGCCTTTTCGAGCAGTAGCTTCTGATAAAAA 2498
DB 185 AGATACCAAAAGATGTGGCAAGGCGCTTGTAGAGCAGTAGCTTCTGATAAAAA 244
2499 TGGAGAGACCTGTTGACCCCTCTCTGACGAGCTCTGTTTACATTCAAGAAAAATGTGT 2558
DB 245 TGGAGAGACCTGTTGACCCCTCTGACGAGCTCTGTTTACATTCAAGAAAAATGTGT 304
2559 CATTCCTGCGCCATTTGATTGCAAGTTAAGCGATTGCTTGTGGGGTCTTTCAGATTTC 2618
DB 305 CATTCCTGCGCCATTTGATTGCAAGTTAAGCGATTGCTTGTGGGGTCTTTCAGATTTC 364
2619 ATCTTGTGGAATTTGGAGTGAATTCGATCAATGCTTAAAGAAAAACCTTACANTGG 2678
DB 365 ATCTTGTGGAATTTGGAGTGAATTTAGATCAATGCTTAAAGAAAAACCTTACANTGG 424
2679 AGGACGACCATGTCCCAAACTGGATCTCAAGAA-----TCAGGTACATGAGGCGAGTCCC 2732
DB 425 AGGTCGCGCCATGTCCCAAACTGGATCTCAAGAACAGGCTCAGGTGCAATGAGGCGAGTCCC 484
2733 ATGTTACAGTGAAGTGAATTCAGTATTCCTGGTGTGTAACACTGCTTTCATGCAAAAAT 2792
DB 485 ATGTTACAGTGAAGTGAATTCAGTATTCCTGGTGTGTAACACTGCTTTCATGCAAAAAT 544
QY 2793 CAACAATGAGTGAAGTCCC 2812
DB 545 CCACATGAGTGAAGTCCC 564

RESULT 13
BM414542 713 bp mRNA linear EST 24-JAN-2002
LOCUS ECC00008 Ing horse testis cDNA Equus caballus cDNA, mRNA sequence.
DEFINITION BM414542
ACCESSION BM414542
VERSION BM414542.1 GI:18314590
KEYWORDS EST.
SOURCE Equus caballus (horse)
ORGANISM Equus caballus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
1 (bases 1 to 713)
Claus,J., Gonzales,R.A., Honeycutt,D., Ing,N., Johnson,L. and
Skow,L.C.
Horse Testis EST Project
Unpublished (2002)
Contact: Skow LC
Dept. of Veterinary Anatomy & Public Health
Texas A&M University
College of Veterinary Medicine, 4458 TAMU, College Station, TX
77843-4458, USA
Tel: 1-979-845-3194
Fax: 1-979-845-9972
Email: lskow@cvm.tamu.edu
FEATURES
JOURNAL
COMMENT
TITLE
JOURNAL
COMMENT

```

Plate: AM001 row: C column: 11

Seq primer: M13 Reverse.

Location/Qualifiers

## FEATURES

source

1..713

/organism="Equus caballus"

/mol\_type="mRNA"

/strain="American Quarter Horse"

/db\_xref="taxon:9796"

/sex="Male"

/lab\_host="XLI-Blue MRF Genotype"

/clone\_lib="ing horse testis cDNA"

/note="Organ: Testis; Vector: Lambda ZAP II Custom cDNA Library; Site 1: EcoR I; cDNA library constructed by Clontech and donated by Nancy Ing, Dept. of Animal Science, Texas A&M University. DNA sequencing by DNA Core Facility, The Noble Foundation, Ardmore, Oklahoma. Clone distribution by L.C. Skow. Send e-mail to: lskow@cvm.tamu.edu."

## ORIGIN

Query Match 10.2%; Score 447.2; DB 12; Length 713;

Best Local Similarity 86.4%; Pred. No. 7.3e-127;

Matches 502; Conservative 0; Mismatches 78; Indels 1; Gaps 1;

QY 1719 GACCAAGAGTGCAGATTCTACTGCACCTGAACCTGTGGCCCTGTTTCTCCCATG 1778

DB 117 GACCAAGAGTGCAGATTCTACTGGCCTGAACAGTAGTCTGTCTCTCCCATG 176

QY 1779 CAAAAAGACTGTATTGTGACTGTTTTCAGTGTGACACCCCTGCCAAGGATGTGCCA 1838

DB 177 CAAAAAGACTGTATTGTGACTGTTTTCAGTGTGACACCCCTGCCAAGGATGTGCCA 236

QY 1839 AGCAGGAATGCCAGTAAACAGTCTCGATACAGATCATCATCCAAAGCAGCCAA 1898

DB 237 ACCAGGAATGCCAGTAAACAGTCTCGATACAGATCATCATCCAAAGCAGCCAA 296

QY 1899 TGGAGGCCAGGAATGCCAGATACCTTATATGAGGAGAGAGTGTGAAGATGTTTCTT 1958

DB 297 CGGAGGCCAGGAGTGCACATACCTTATTTGAAGAGCGAATGTGAAGATGTTCTATT 356

QY 1959 GTGTCTCTATATCGTGGAGGCCACAGAAATGAGAGCCCTTGCATCTTAGTCCAGATC 2018

DB 357 GTGCCCATATATCGTGGAGGCTGCAGAAATGAGAGCCCTTGTTCCTGTCAGAGTC 416

QY 2019 TGTCTGGCAGGAATPACGGCAGCAGTGAACCTGTGGAAGGGGTACAAACAGAGC 2078

DB 417 COTCAGGAAGGAATATATGGGCACCATGTGAAGCCTGTGGAAGGGGTACAAACAGAGC 476

QY 2079 TGTCTCATGTCATCTCTGATGACAAACCGTTCAGCAGAAATGATGGAATGCTCAAGCAGAC 2138

DB 477 TGTCTCATGTCATCTCTGATGACATCCAGTCAGCAGAAATGACAGATGTTCTCAAGCAGAC 536

QY 2139 AAACGGCATGCTCTCTGTCAGAGATGCACAGTCCCATGTGCGAGAGACTGCACCTT 2198

DB 537 AAATGGCATGCTCTCTGTCAGAGATGCACAGTCCCATGTGCGAGAGACTGCACCTT 596

QY 2199 CACTGCTTGGTCCCAAGTTTACGCCCTGTCCACGAACTGTGAAGCCACAAAGATAGCG 2258

DB 597 CACTGCTTGGTCCCAAGTTTACACCTTGTCTGCAAAATTGNGAA-CAACCCGAGTANAGC 655

QY 2259 GCGAGCTCAGGGAAAGCAGAAAGAGAGAAATGCC 2299

DB 656 GCGAGCTCAGAGTTATGTCAANAANAATGCNTATCCC 696

## RESULT 14

BE899999

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

BE899999 551 bp mRNA linear EST 25-APR-2001

201865 MARC 450V Bos taurus cDNA 5', mRNA sequence.

BE899999

BE899999.1 GI:10387839

EST.

Bos taurus (cow)

## ORGANISM

Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.

## REFERENCE

AUTHORS

1 (bases 1 to 551)

Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T., Casas, E., Wray, J.E., White, J., Cho, J., Fahrner, S.C., Bennett, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G., Pette, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and Keefe, J.W.

Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle

Genome Res. 11 (4), 626-630 (2001)

21180013

11182978

CONTACT: Smith TPL

USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366

Fax: 402 762 4390

Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called and alt trimmed with phred

v0.980904.e. Vector identified by cross\_match with the -minscore 18

and -mismatch 12 options.

PCR Primers

FORWARD: AGGAAACAGCTATGACCAT

BACKWARD: GTTTCCTCAGTCAGACG

Plate: 112 row: C column: 8

Seq primer: ATTAGTGACACTATAG.

Location/Qualifiers

1..551

/organism="Bos taurus"

/mol\_type="mRNA"

/db\_xref="taxon:9913"

/tissue\_type="pooled"

/lab\_host="DH10B"

/clone\_lib="MARC 450V"

/note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;

Library made from pooled tissue from day 20 and day 40

embryos."

embryos."

## ORIGIN

Query Match 10.1%; Score 446; DB 10; Length 551;

Best Local Similarity 88.2%; Pred. No. 1.4e-126;

Matches 485; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 1550 ATTCTGTATGAGCTTCACTGGGAGACATCGCCTTGGGGCCCTTGTTCGAGACACAT 1609

DB 1 ATTCTGTATGAGCTTCACTGGGAGACATCACCTTGGGTCCCTGTTCTGAAGACACAT 60

QY 1610 TGTAACTCCCTTAATGCAACATTTGGCTGGAATGGAGAGCCACGTTGTTGTAGGCA 1669

DB 61 TAGTAATCGCCTTAATGCAACATTTGGCTGGAATGGAGAGCAACATGTTGTTGGGCA 120

QY 1670 TTTCAGACTCGGAGAGTCTTCTGTCAAGAGTCAGTGGCAAGCAATGACCAAAAGAT 1729

DB 121 TTTCAGACTCGGAGAGTCTTCTGTCAAGAGTCAGTGGCAAGCAATGACCAAAAGAT 180

QY 1730 GTCCAGATTTACTGACCTGAAACCTGTGCGCCCTTGTTCCTCCATGCAAAAAGACT 1789

DB 181 GTCCAGATTTACTGCGCCAGAAACCTGTACGGCTTGTCTGCCCTGCAAAAAGACT 240

QY 1790 GTATTGTGACTGTTTCACTGAGTGGACACCTTGCCTCAGGATGTGCCAGGAGAAATG 1849

DB 241 GTCTCGTGAATGCTTTTCACTGAGTGGACTCCCTTGCCTCAGGCTGTGCCAAGGAAATG 300

QY 1850 CCACAGTAAACAGTCTCCATACAGATCATCATCAAGAGCAGCCCAATGAGGCCAGG 1909

DB 301 CTACATTAACAGTCTCCATACAGATCATCATCAAGAGCAGCCCAATGAGGCCAGG 360

QY 1910 AATGCCAGATACCTTTATATGAGGAGAGAGATGTGAAGATGTTTCTTCTCTCTGAT 1969

DB 361 AATGCCAGATACCTTTATTTGAAGAGCAGAGATGTGAAGATGTTTCTTCTCTCTGAT 420

Y 1970 ATCGGTGGAGCCACAGAAATGGAGCCCTTGCATCTTAGTCCAGAGTCTCTGGCAGG 2029  
b 421 ATCGATGGAGGCCACAGAAATGGAGCCCTTGTCTCGTCCAGAGTCTCTCGGCAGG 480  
Y 2030 GAATAACGGGACGAGTGAAGCCTGTGGAAGGGGTTACAAACAGAGAGCTGTCTCATGCA 2089  
b 481 GACTAATAGCCACCACTGTAACCTCGGGAAGGGAGTGCAATAAGAGCTGTCTCATGCA 540  
Y 2090 TCTCTGATCA 2099  
b 541 TCTCTGATAA 550

RESULT 15  
LOCUS BI991468 600 bp mRNA linear EST 20-DEC-2001  
DEFINITION 4101-38 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA,  
mRNA sequence.  
ACCESSION BI991468  
VERSION BI991468.1 GI:17962491  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
REFERENCE 1 (bases 1 to 600)  
AUTHORS Mu,X., Zhao,S., Pershad,R., Hsieh,T.-P., Scarpa,A., Wang,S.W.,  
White,R.A., Beremand,P.D., Thomas,T.L., Gan,L. and Klein,W.H.  
TITLE Gene expression in the developing mouse retina by EST sequencing  
and microarray analysis  
JOURNAL Nucleic Acids Res. 29 (24), 4983-4993 (2001)  
MEDLINE 21671825  
PUBMED 11812828  
COMMENT Contact: Klein WH  
Department of Biochemistry and Molecular Biology  
University of Texas M.D. Anderson Cancer Center  
Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA  
Tel: 713 792 3646  
Fax: 713 790 0329.  
Location/Qualifiers  
1. .600  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/tissue\_type="neural retina"  
/dev\_stage="embryonic day 14.5 post-fertilization"  
/clone\_lib="Mouse E14.5 retina lambda:ZAP II Library"

FEATURES  
source

Query Match 9.6%; Score 422.4; DB 12; Length 600;  
Best Local Similarity 82.7%; Pred. No. 3.5e-119;  
Matches 496; Conservative 0; Mismatches 101; Indels 3; Gaps 1;  
Y 1882 ATCAAGAAGCAGCAATGAGGCGCAGGAATGCCAGATACCTTTATATAGGAGAGAG 1941  
b 1 ATCCAGGAGCAGCCATGGGGTCAAGAGTGCCAGACACCTTTGTTGAAGAAAGAGAG 60  
Y 1942 TGTGAAGATGTTCTTGTGTCCTGTATATCGGTGGAGCCACAGAAATGGAGCCCTTGC 2001  
b 61 TGTGAAGATATCTCACTGTGTCTTATACCGGTGGAGCCGAGAAATGGAGCTTTGT 120  
Y 2002 ATCTTAGTGCCAGAGTCTGTGCGCGGAATACGGGCGAGCTGTAAGCCTGTGGAAG 2061  
b 121 ATCTGTGTTCCAGAGTATATCAGCAGGGGAGAACAGCCACCATGAGGATGTGGGAG 180  
Y 2062 GGGTTACAAACAGAGGTGTCATGATCTCTGATGACACCGGTGAGCAAAATGATG 2121  
b 181 GGGCTGCAGACAAGAGTGTCTCATGATCTCTGATGATTAACCATCTGTCAGAAATGACA 240  
Y 2122 GAATGCTCTCAGACAGACAACCGCATGCCTCTCTGTGCAAGATGCACAGTCCCATGT 2181  
b 241 GAGTGCCTCAAGCAGATGATGGCATGCTCCCTCTTGTGAGGAATGACATCTCCCTGT 300



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

DM nucleic - nucleic search, using sw model

run on: February 23, 2004, 12:03:01 ; Search time 329.257 Seconds  
(without alignments)  
7412.668 Million cell updates/sec

Title: US-10-022-710-1  
Perfect score: 4398  
Sequence: 1 atgggtgcgtgattcagaa.....atggagacttagacatgtaa 4398

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq: \*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq: \*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq: \*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq: \*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq: \*  
6: /cgn2\_6/ptodata/2/ina/backfileseq1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1229.6	28.0	6378	4	US-09-620-312D-332
2	161.8	3.7	579	4	Sequence 332, Appl
3	44	1.0	3226	1	US-07-862-021B-11
4	44	1.0	3226	1	US-08-313-288B-11
5	44	1.0	3226	5	PCT-US93-03164-11
6	42.8	1.0	7218	1	US-08-232-463-14
7	39.8	0.9	4061	4	US-09-132-769-4
8	39.2	0.9	3079	4	US-09-132-769-2
9	39.2	0.9	3079	4	US-09-640-173-184
10	39.2	0.9	3079	4	US-09-713-550-184
11	39.2	0.9	3999	4	US-09-640-173-177
12	39.2	0.9	3999	4	US-09-713-550-177
13	39.2	0.9	4866	4	US-09-976-594-353
14	37	0.8	505	4	US-09-621-976-15639
15	37	0.8	576	4	US-09-397-992A-25
16	37	0.8	576	4	US-09-971-843-25
17	37	0.8	832	4	US-09-621-976-2813
18	36.6	0.8	832	4	US-09-621-976-2813
19	36.6	0.8	4029	1	US-07-862-021B-9
20	36.6	0.8	4029	1	US-08-313-288B-9
21	36.6	0.8	4029	4	US-09-132-769-6
22	36.6	0.8	4029	5	PCT-US93-03164-9
23	36.6	0.8	882	4	US-09-252-991A-2427
24	36.2	0.8	993	4	US-09-252-991A-2588
25	35.8	0.8	474	4	US-09-621-976-18033
26	35	0.8	1668	4	US-09-252-991A-10888
27	35	0.8	1848	4	US-09-252-991A-10829

Sequence 2427, Ap  
Sequence 3, Appli  
Sequence 15, Appl  
Sequence 19, Appl  
Sequence 5, Appli  
Sequence 165, App  
Sequence 298, App  
Sequence 50, Appl  
Sequence 67, Appl  
Sequence 67, Appl  
Sequence 67, Appl  
Sequence 67, Appl  
Sequence 7862, Ap  
Sequence 7895, Ap  
Sequence 7515, Ap  
Sequence 7823, Ap  
Sequence 8976, Ap

ALIGNMENTS

RESULT 1  
US-09-620-312D-332  
; Sequence 332 Application US/09620312D  
; Patent No. 6569662  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Ren, Feiyan  
; APPLICANT: Chen, Rui-hong  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Yang, Yonghong  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Zhou, Ping  
; APPLICANT: Ma, Yundong  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Wang, Zhiwei  
; APPLICANT: John Tillinghast  
; APPLICANT: Dmanac, Radoje T.  
; TITLE OF INVENTION: No. 5569662el Nucleic Acids and  
; TITLE OF INVENTION: Polypeptides  
; FILE REFERENCE: 784CIP2B  
; CURRENT APPLICATION NUMBER: US/09/620,312D  
; CURRENT FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 1105  
; SOFTWARE: pt FL\_genes Version 1.0  
; SEQ ID NO 332  
; LENGTH: 6378  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (271)..(5037)  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(6378)  
; OTHER INFORMATION: n = a,t,c or g  
US-09-620-312D-332

Query Match 28.0%; Score 1229.6; DB 4; Length 6378;  
Best Local Similarity 57.6%; Pred.No. 0;  
Matches 2596; Conservative 0; Mismatches 1784; Indels 126; Gaps 17;



2692 GGAGGACAGGCTGGAATCCATGAGTGCCCTACAGTATGACAGGCCCTGTGCGAGCCCTTACC 2751  
2161 CAAGAATGACAGTCCCATGTGCGAAGACTGACCTTCACTGCTTGTGTCACAGTTTACG 2220  
2752 CAGGCTGCGCAGATCCCTGCCAGGATGACTGTCAATTTACACAGCTGGTCCAAGTTTCT 2811  
2221 CCTGTCTCCAGCAACTGTGAAGCCACAAAGTGTAGCGGCGACAGCTCACAGGGAAGC 2280  
2812 TCATGCAATGGAGACTGTGGTGCAGTTAGGACACAGAAAGCGCACTCTTGTGGAAGAGT 2871  
2281 AGAAGAAGAGGAATGCCAGGATTTGACCTTTACCTCTTAGTGGAGACAGAACTATGT 2340  
2872 AAAAAAAAAAAAAATGTAATAATTCCTATTTGTATCCCTGATTTGAGACTCAGTATTGT 2931  
2341 CCTGTGATGAATTTATATCCCAACCTTATGAAACTGTGAACTGTGCAATTCCTCCAGAA 2400  
2932 CTTTGTGACAAATATATGACAACTGTGGGAACTGTGACAGTGTATTTTACCAGAG 2991  
2401 GGCAGAAAGGAGCCTCACCGAGACTGCGGGTACAAGCAGACAGAAAGAAATGTGGAGAA 2460  
2992 GGAAGAGTGGAGTGTGCTGGGAATGAAAGTACAAGGAGACATCAAGGAATGCGGACAA 3051  
2461 GGCCTGGCTTTCGACAGTGTGCTGCTGATGAAATGGAAGACCTGTTCACCCCTCC 2520  
3052 GATATCGTTACCAAGCAATGGCATCTAGCATCAAAATGGCAGGCTTGTGGAACAICT 3111  
2521 TTCTGCAGCAGCTCTGGTTACATTCAGAAAAATGTGTCATTCCTGCCCATTTGATTGC 2580  
3112 AGATGTAACAGCCATGTTTACATTTAGGAGGCGCTGCATCATCCCTGCCCTCAGACTGC 3171  
2581 AAGTTAAGCAATGGTCTAGTTGGGGCTTTCAGTTTCACTTGTGGATTTGGAGTGAGA 2640  
3172 AAGCTCAGTGAGTGGTCCAACTGGTGGCTGAGCAAGTCTGTGGGAGTGGTGTGAAG 3231  
2641 ATTTCGATCCAAATGGCTTAAAGAAAAACCTTACAATGGAGGACGACCATGTCCCAACTG 2700  
3232 GTTCGTTCTAAATGGCTGCGTGAAACCATATAATGGAGGAGGCCCTTGCCCAACTG 3291  
2701 GATC-----TCAGAAATCAGGTACATGAGGAGTCCCATGTTACGTGAGTGCAATCAG 2754  
3292 GACCATGTCAACAGGACAGAGTGTATGAGTTGTCCCATGCCACAGTGACTGCAACCCAG 3351  
2755 TATTCTGGGTTGTAGAACACTGGTCTTCATGCAAAATCAACAATCAGCTGAGTGCCTTG 2814  
3352 TACTATGGGTACAGAGCCCTGAGCACTGCAAGGTGACCTTTGTGAATATGCGGAG 3411  
2815 CGCTGTGGAGGAGAAACAATCTAGGAAATCAGATGTGT---GAATATCGGATGT 2871  
3412 AACTGTGGAGAGGCGGTGCAAACCCGAAAGTGAATGATGCATGCAGAAATACAGCAGATGC 3471  
2872 GAAGTGGAGCAGTGGATAGCAACCTGTGCAACAGGATGMAATTCCTCCAGAAACCCAG 2931  
3472 CCTTCTGAACATGTAGAGGATTTACTCTGTGACCCAGAGAGATGCCCTGGCTCTAGA 3531  
2932 TCCTGTTCTCTTATGTCTCCATGATGTGTCTATGTCTGAGTGGGAGCTTTGGAGCAAA 2991  
3532 GTGTGCAAAATTAACATCCCTGAGGACTGTGTGATATCTGAATGGGCTCCATGGACCCAA 3591  
2992 TGCCCAAGTCACTGCGATCCCAACAAAT---GCAGAGAAAGCTGCCACCTGCTTAAGA 3048  
3592 TGTGTTTTGGCTTGCATCAAGAGCAGTTTCCGCAAAAGGTGACGTGATCCCATCAGACAA 3651  
3049 CCATCACTGAACTCAGGACTTTGTCTGGAAGACTCAGAGTGGAGCTTGCCTCTCTGAAT 3108  
3652 CCAGCTGATGAGGAAGATCTTGCCCTAATGCTTTGAGAAAGAACCCGTAACTCTGAAC 3711  
3109 GAAATTTGCTTCCAGTTCAGTACAATCTAACAGAGTGGAGCACTGCCAGCTGAGTGAA 3168  
3712 AAAAACTGCTACCACTATGATTAATGTAACAGACTGGAGTACATGTACAGCTGAGTGAG 3771  
3169 AACCCACCTGTGTGCAAGGCTCAGAGCCGCTGCTAAAGCTGTGTGTCAGTGTATGCC 3228  
3772 AAGCGAGTTTGTGGAATGGAATATAAACAAGGATGTGGATTTGATTTGCAAGTGTATGCC 3931

QY 3229 AAGCAGTCAAGCTGACCAATGTGAGCAGCATAAATTTGGAGAAAGCCCCAGAGAAATGAGC 3288  
Db 3832 AAGTCAGTTCACTGAAATAATTTGTGAAGCGCTTGGCTTTGGAGAGAACTGCGCAGATGAC 3891  
QY 3289 ATTCCCTGCTTGTGGAATGCGTGGTCACTGTCACTGTCAAGGTGAGCGGTGGACAA 3348  
Db 3892 ACGTCTGCATGTGTGAATGCCCTGTGAACATGTCACTGTCTGATTTGCTCTCTGCTCA 3951  
QY 3349 GAGTGTTCACAGACCTGTGGCCATGAGGTGCAATGAGCCGAGCTCGATTTATCATTTATG 3408  
Db 3952 GAATGTTCTCAACATGTGCTTCAAGGAAAAATGATCCGAAGACGACAGTGAACCCAG 4011  
QY 3409 CCACCCCAAGGAGAGGACGCGCCATGCCCAAGAGCTTACCCAGAGAAAAAOCCTGCCCA 3468  
Db 4012 CCCTTTCAAGGTGATGGAAGACCATGCCCTTCCCTGATGAGCCAGTCCAAACCCCTGCCCA 4071  
QY 3469 GTGACCCCTCTCTACAGCTGGTCTTGGCAATGTGTCTGCATGTAAATTTGGAGGGTGA 3528  
Db 4072 GTGAAGCTTTTATCGGTGGCAATATGCGCAGTGGTCTCCATGCCAAGTGCAGAGGCC 4131  
QY 3529 GACTGTGGGGAAGAGTTTCAGATCCGAGCGCTTTCTCTGCATGGTCCACAGTGGTTCAATA 3588  
Db 4132 CAGTGTGAGAGAGGAGACCAAGAAACATTTCTTGTGTAGTAAAGTATGGTCACT 4191  
QY 3589 TCTCATGAGCTGGAAGTGTGAGGATGCACTGTGTGGAGAAATGCCCC----- 3638  
Db 4192 GATGATTTCAAGAAAGTGTGGATGAGGAATTTCTGTCTGACATTTGAACCTCATTTATGAT 4251  
QY 3639 -----TCAGGACACATCTCTGAAGCAGCTGTGTCTGTGCTTGGCCAGGAGCTGCCAT 3693  
Db 4252 GGTAAATAAATAATGTTCTTGGAGGAATCCTGCAGCAGCCCTTGGCCAGTGTAT 4311  
QY 3694 TTAACAGAAATGGTCAGAGTGGAGCACAATGTGAATTAACCTGCAATGTATGGAAGAGCTTT 3753  
Db 4312 TTGAAGGACTGGTCTTCTGAGGCTGTGTGTCAGTGAACCTGTGTGAATGTGTGAGATCTA 4371  
QY 3754 GAGACTGTGGGCCCGCAGTCTAGATCAAGGACTTTTATAATTCAGTCTTTTGAAGAACAA 3813  
Db 4372 GCTTTGTGGAATACAGGTCAAGTCCAGACCGGTGATTTACAGAACTAGAGANTCAG 4431  
QY 3814 GACAGCTGCCCCCAACAGGTTCTAGAAACAGCCCTTGTACAGGAGCAAAATGTTATCAC 3873  
Db 4432 CATCTGTGCCCAGAGCAGATGTTAGAAACAAATCATGTTATGATGACACAGTGTATGAA 4491  
QY 3874 TACACATGGAAGCAAGTCTTTGGAACAATAACGAACTGTATGTCGCCAGCTTCA 3933  
Db 4492 TATAATGGAATGGCCAGTGTGTGGAAGGCTCTTCCGAAACAGTGTGTGTCAAGTCA 4551  
QY 3934 GATGGCGTTAATGTCAAGAGGAGTGTCTCCCTCAGGCCCGTCTCTGCTGCCATTCGGCAG 3993  
Db 4552 GATGTTAATGTAACAGGGGGCTGCTTGGTGTATGAGCCAGCCTGATGCCAGAGTCT 4611  
QY 3994 TGCATTCAGCTGCAAGAAACCTTTCTCTACTGTACACAGGGTGGAGTCTGTGGTGT 4053  
Db 4612 TGTAAACCCACCGTGTAGTCAACCCCTCTGCTACTGTAGCAGACAAACATGCCATGT 4671  
QY 4054 GAGAAGGCTATACAGAGATAATGAAATCAAAATGTTTCTCTGGAATTAAGTCA----- 4105  
Db 4672 GAAGAGGGTACACTGAGTCACTCTTTTAAACAGCACCCTTGAGCAATGCACACTTATC 4731  
QY 4106 -----TGAAAGTACCAAGGCTCAGAGATAAAGAGTGTGTAAGAAACCTTTCTG-G 4157  
Db 4732 CCGCTGGTGTGTTTACCACCATGGAGACAAAAGAGGAGATGTGAAACCCAGTCGGCT 4791  
QY 4158 GAAAGACAGCTGTGAATTTCAAAATACATGATATTTTTAAAGGA--TGCTCTCTCAA 4215  
Db 4792 GTACATCCAAACCCACCTTCCAGTAAACAGAGACGGGGAAGGACCTGTGTTCTACAG 4851  
QY 4216 CCATTTGATCCAGATGGCCGAGTAAAAATTTGGGTTTATGGCGTTTTCAGTGGCGCTTTT 4275  
Db 4852 CCATTTGGCCAGATGGGAGACTAAAGACCTGGGTTTACGGGTGTAGCAGCTGGGCGCATTT 4911



IS-08-313-288B-11  
Sequence 11, Application US/08313288B  
Patent No. 5750502

## GENERAL INFORMATION:

APPLICANT: Jessell, Thomas M. and AviHu Klar  
TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A  
TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/313,288B  
FILING DATE: January 5, 1995  
CLASSIFICATION: 435

## ATTORNEY/AGENT INFORMATION:

NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 40028-A-PCT-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0526  
TELEX:

INFORMATION FOR SEQ ID NO: 11:

## SEQUENCE CHARACTERISTICS:

LENGTH: 3226 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:

NAME/KEY: CDS

LOCATION: 136..2543

US-08-313-288B-11

Query Match 1.0%; Score 44; DB 1; Length 3226;  
Best Local Similarity 58.3%; Pred. No. 0.0047;  
Matches 77; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 2566 TGCCATTTCGACGATTGCTAGTTGGGGGCTTCGACGTTCACTTGT 2625  
DB 2113 TGCCCATTTGACTGTGAGCTAACGGAGTGGTCCGAGTGCAATCCTCTGT 2172

QY 2626 GGAATTGGAGTGAGATTCGATCCAAATGCTAAAGAAAACCTTACAATGGAGGACGA 2685  
DB 2173 GGGAGGGCCACATGATCAGAACAAAGATGATCAAAATAGACACACAGTTTGGAGGAACA 2232

QY 2686 CCATGTCCTCCAAA 2697

DB 2233 GCATGCCCAGNA 2244

## RESULT 5

PCT-US93-03164-11

Sequence 11, Application PC/TUS9303164

## GENERAL INFORMATION:

APPLICANT: Jessell, Thomas M  
APPLICANT: Klar, AviHu  
TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A  
TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham  
STREET: 30 Rockefeller Plaza

CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10112  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/03164  
FILING DATE: 19930402  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 40028  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 977-9550  
TELEFAX: (212) 664-0525  
TELEX: 422523 COOP UI  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3226 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:

NAME/KEY: CDS

LOCATION: 136..2543

PCT-US93-03164-11

Query Match 1.0%; Score 44; DB 5; Length 3226;

Best Local Similarity 58.3%; Pred. No. 0.0047;

Matches 77; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 2566 TGCCATTTCGACGATTGCTAGTTGGGGGCTTCGACGTTCACTTGT 2625  
DB 2113 TGCCCATTTGACTGTGAGCTAACGGAGTGGTCCGAGTGCAATCCTCTGT 2172

QY 2626 GGAATTGGAGTGAGATTCGATCCAAATGCTAAAGAAAACCTTACAATGGAGGACGA 2685  
DB 2173 GGGAGGGCCACATGATCAGAACAAAGATGATCAAAATAGACACACAGTTTGGAGGAACA 2232

QY 2686 CCATGTCCTCCAAA 2697

DB 2233 GCATGCCCAGAA 2244

## RESULT 6

US-08-232-463-14/c

Sequence 14, Application US/08232463

Patent No. 5670367

## GENERAL INFORMATION:

APPLICANT: DORNER, F.  
APPLICANT: SCHEIFELINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION NUMBER: US/08/232.463  
FILING DATE: 26-AUG-1991  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935.313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZgpt-F1s  
US-08-232-463-14

Query Match 1.0%; Score 42.8; DB 1; Length 7218;  
Best Local Similarity 4.3%; Pred. No. 0.022;  
Matches 17; Conservative 210; Mismatches 167; Indels 0; Gaps 0;  
QY 1780 AAAAAGACTGATGTCGCTTTTCAGTGAGTGACACCTGCCCAAGGATGCGCAA 1859  
DB 1426 RRR 1367  
QY 1840 GCAGGAATGCCAGTAAACAGTCTCGATACAGAAATCATCAAGAGAGAGCAAT 1899  
DB 1366 RRR 1307  
QY 1900 GGAGCCAGGAATGCCAGATACCTTATGAGGAGAGAGAGTGGAAGTCTCTTG 1959  
DB 1306 RRR 1247  
QY 1960 TGTCTGTATATCGGTGGAAGCACAGAAATGGAGCCCTTCATCTTAGTCCAGAGTCT 2019  
DB 1246 RRR 1187  
QY 2020 GTCTGCAGGAATTAACGGGACAGTGAAGCTTGGAAGGGTTACAAACAGAGCT 2079  
DB 1186 RRR 1127  
QY 2080 GTCTCATGCTCTCTGATGACAAACCGGTGACAGAAATGATGGAATGCTCAAGAGACA 2139  
DB 1126 RRR 1067  
QY 2140 AACGGGATGCTCTCTGTGTGAAGAAATGCAAG 2173  
DB 1066 ATCGAAGCTCTCGACCTGACGCAAGCTCGG 1033

RESULT 7  
US-09-132-769-4  
Sequence 4, Application US/09132769A  
Patent No. 6525023  
GENERAL INFORMATION:  
APPLICANT: Motoco Yamasaki  
APPLICANT: Kenji Shibata  
APPLICANT: No. 6525023uo Hanai  
APPLICANT: Kaoru Miyamoto  
TITLE OF INVENTION: NOVEL VASCULAR SMOOTH MUSCLE CELL GROWTH FACTOR

FILE REFERENCE: 11078  
CURRENT APPLICATION NUMBER: US/09/132,769A  
CURRENT FILING DATE: 1998-08-12  
EARLIER APPLICATION NUMBER: HEI9-218491  
EARLIER FILING DATE: 1997-08-13  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 4  
LENGTH: 4061  
TYPE: DNA  
ORGANISM: BOVINE  
US-09-132-769-4

Query Match 0.9%; Score 39.8; DB 4; Length 4061;  
Best Local Similarity 56.5%; Pred. No. 0.14; 57; Indels 0; Gaps 0;  
Matches 74; Conservative 0; Mismatches 0;  
QY 103 TGTCCCGGGATTGTGTAGTATCTGAGTTCTTACCATGTCCTCAACTGTAGCAAGGATGT 162  
DB 2119 TGCCCCATTGACTGTGAGCTCAGCGAGTGGTCCAGTGGTGGAAATGTAAACAAGTCAATG 2178  
QY 163 GGGAGAAATTCAGCATAGAACTCGCGGGTCATAGCTCCCCCTCTCTTTGGTGGTTG 222  
DB 2179 GGGAAAGGACATCATGATTGGAACCCCGCATGATCCAAATGGAGCCTCAGTTTGGAGGCACA 2238  
QY 223 CAATGTCCAAA 233  
DB 2239 CCTGCCCCAGA 2249

RESULT 8  
US-09-132-769-2  
Sequence 2, Application US/09132769A  
Patent No. 6525023  
GENERAL INFORMATION:  
APPLICANT: Motoco Yamasaki  
APPLICANT: Kenji Shibata  
APPLICANT: No. 6525023uo Hanai  
APPLICANT: Akiko Furuya  
APPLICANT: Kaoru Miyamoto  
TITLE OF INVENTION: NOVEL VASCULAR SMOOTH MUSCLE CELL GROWTH FACTOR  
FILE REFERENCE: 11078  
CURRENT APPLICATION NUMBER: US/09/132,769A  
CURRENT FILING DATE: 1998-08-12  
EARLIER APPLICATION NUMBER: HEI9-218491  
EARLIER FILING DATE: 1997-08-13  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 3079  
TYPE: DNA  
ORGANISM: human  
US-09-132-769-2

Query Match 0.9%; Score 39.2; DB 4; Length 3079;  
Best Local Similarity 50.5%; Pred. No. 0.17; 93; Indels 0; Gaps 0;  
Matches 95; Conservative 0; Mismatches 0;  
QY 46 GAAATATGCGAACACATTTGCGCTTCAGCCTCCTACAGAACAGGCTTGCTCATTCTTGT 105  
DB 1960 GAATTTGGAGACTGCAATGAGGATCTGGAGAGGTGGAGAAAGTGCATGCTCCTGAATGC 2019  
QY 106 CCCCCGGATTGTGTAGTATCTGAGTTCTTACCATGTCCTCAACTGTAGCAAGGATGTGGG 165  
DB 2020 CCCATTGACTGTGAGCTCACCGAGTGGTCCCAAGTGGTGGAAATGTAAACAAGTCAATG 2079  
QY 166 AAGAAATTCAGCATAGAACTCGCGGGTCATAGCTCCCCCTCTCTTTGGTGGTTGCAA 225  
DB 2080 AAAGGCCAGTGTATTCGAACCCCGCATGATCCAAATGGAGCCTCAGTTTGGAGGTGCA 2139  
QY 226 TGTCCTCAA 233  
DB 2140 TGCCCCAGA 2147



```

; RESULT 9
; US-09-640-173-184
; Sequence 184, Application US/09640173
; Patent No. 6617109
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: OVARIAN TUMOR SEQUENCES AND
; TITLE OF INVENTION: METHODS OF USE THEREFOR
; FILE REFERENCE: 210121.484C2
; CURRENT APPLICATION NUMBER: US/09/640,173
; CURRENT FILING DATE: 2000-08-15
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 184
; LENGTH: 3079
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-640-173-184

Query Match      0.9%; Score 39.2; DB 4; Length 3079;
Best Local Similarity 50.5%; Pred. No. 0.17; 93; Indels 0; Gaps 0;
Matches 95; Conservative 0; Mismatches 0;

2Y 46 GAAATATGCGAACAACCTTGGCCCTTCAGCTCCTACAGAACAGGCTTGCCTCATTCCTTGT 105
Db 1960 GAACCTTGAGACTGCAATGAGGATCTGGAGCAGGTGGAGAGTGCATGCTCCCTGAATGC 2019
2Y 106 CCCGGGATGTGTAGTATCTGAGTCTTACCAGTGTCCAACTGTAGCAAGGATGTGG 165
Db 2020 CCATTGACTGTGAGCTACCGAGTGTGCCAGTGTGCCAGTGTGGAATGTAAAGATCATGTGG 2079
2Y 166 AAGAAATTGCGACATAGAACTCCGGGGTGCATAGCTCCCTCTCTTTGGTGGTTTGCAA 225
Db 2080 AAGGCCACGTGATTGCAACCCGGATGATCCAAATGGAGGCTCAGTTTGAGGTGCACCC 2139
2Y 226 TGTCCAAA 233
Db 2140 TGCCACAGA 2147

; RESULT 10
; US-09-713-550-184
; Sequence 184, Application US/09713550
; Patent No. 6617109
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.484C4
; CURRENT APPLICATION NUMBER: US/09/713,550
; CURRENT FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 205
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 184
; LENGTH: 3079
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-713-550-184

Query Match      0.9%; Score 39.2; DB 4; Length 3079;
Best Local Similarity 50.5%; Pred. No. 0.17; 93; Indels 0; Gaps 0;
Matches 95; Conservative 0; Mismatches 0;

2Y 46 GAAATATGCGAACAACCTTGGCCCTTCAGCTCCTACAGAACAGGCTTGCCTCATTCCTTGT 105
Db 1960 GAACCTTGAGACTGCAATGAGGATCTGGAGCAGGTGGAGAGTGCATGCTCCCTGAATGC 2019
2Y 106 CCCGGGATGTGTAGTATCTGAGTCTTACCAGTGTCCAACTGTAGCAAGGATGTGG 165
Db 2020 CCATTGACTGTGAGCTACCGAGTGTGCCAGTGTGCCAGTGTGGAATGTAAAGATCATGTGG 2079
2Y 166 AAGAAATTGCGACATAGAACTCCGGGGTGCATAGCTCCCTCTCTTTGGTGGTTTGCAA 225
Db 2080 AAGGCCACGTGATTGCAACCCGGATGATCCAAATGGAGGCTCAGTTTGAGGTGCACCC 2139
2Y 226 TGTCCAAA 233
Db 2140 TGCCACAGA 2147

; RESULT 11
; US-09-640-173-177
; Sequence 177, Application US/09640173
; Patent No. 6613515
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: OVARIAN TUMOR SEQUENCES AND
; TITLE OF INVENTION: METHODS OF USE THEREFOR
; FILE REFERENCE: 210121.484C2
; CURRENT APPLICATION NUMBER: US/09/640,173
; CURRENT FILING DATE: 2000-08-15
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 177
; LENGTH: 3999
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-640-173-177

Query Match      0.9%; Score 39.2; DB 4; Length 3999;
Best Local Similarity 50.5%; Pred. No. 0.21; 93; Indels 0; Gaps 0;
Matches 95; Conservative 0; Mismatches 0;

2Y 46 GAAATATGCGAACAACCTTGGCCCTTCAGCTCCTACAGAACAGGCTTGCCTCATTCCTTGT 105
Db 1387 GAACCTTGAGACTGCAATGAGGATCTGGAGCAGGTGGAGAGTGCATGCTCCCTGAATGC 1446
2Y 106 CCCGGGATGTGTAGTATCTGAGTCTTACCAGTGTCCAACTGTAGCAAGGATGTGG 165
Db 1447 CCATTGACTGTGAGCTACCGAGTGTGCCAGTGTGCCAGTGTGGAATGTAAAGATCATGTGG 1506
2Y 166 AAGAAATTGCGACATAGAACTCCGGGGTGCATAGCTCCCTCTCTTTGGTGGTTTGCAA 225
Db 1507 AAGGCCACGTGATTGCAACCCGGATGATCCAAATGGAGGCTCAGTTTGAGGTGCACCC 1566
2Y 226 TGTCCAAA 233
Db 1567 TGCCACAGA 1574

; RESULT 12
; US-09-713-550-177
; Sequence 177, Application US/09713550
; Patent No. 6617109
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.484C4
; CURRENT APPLICATION NUMBER: US/09/713,550
; CURRENT FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 205
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 177
; LENGTH: 3999
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-713-550-177

Query Match      0.9%; Score 39.2; DB 4; Length 3999;
Best Local Similarity 50.5%; Pred. No. 0.21; 93; Indels 0; Gaps 0;
Matches 95; Conservative 0; Mismatches 0;

2Y 46 GAAATATGCGAACAACCTTGGCCCTTCAGCTCCTACAGAACAGGCTTGCCTCATTCCTTGT 105
Db 1960 GAACCTTGAGACTGCAATGAGGATCTGGAGCAGGTGGAGAGTGCATGCTCCCTGAATGC 2019
2Y 106 CCCGGGATGTGTAGTATCTGAGTCTTACCAGTGTCCAACTGTAGCAAGGATGTGG 165
Db 2020 CCATTGACTGTGAGCTACCGAGTGTGCCAGTGTGCCAGTGTGGAATGTAAAGATCATGTGG 1506
2Y 166 AAGAAATTGCGACATAGAACTCCGGGGTGCATAGCTCCCTCTCTTTGGTGGTTTGCAA 225
Db 1507 AAGGCCACGTGATTGCAACCCGGATGATCCAAATGGAGGCTCAGTTTGAGGTGCACCC 1566
2Y 226 TGTCCAAA 233
Db 1567 TGCCACAGA 1574
```

Best Local Similarity 50.5%; Pred. No. 0.21;  
Matches 95; Conservative 0; Mismatches 93; Indels 0; Gaps 0;  
QY 46 GAATATGCGAACACTTTGCCCTTCAGCCTCTCTACAGAACAGGCTTGCTCTATTCCTTGT 105  
DB 1387 GAACTTGGAGAGTGCATGAGAGTCTGGAGCAGGTGGAGAGTGCATGCTCTCTGAATGC 1446  
QY 106 CCCCGGATTTGTAGTATCTGAGTTCTTACCATGCTCCAACTGTAGCAAGGATGTGGG 165  
DB 1447 CCCATTGACTGTGAGCTCACCGAGTGTCCAGTGTGCGAATGTACAACTCATGTGGG 1506  
QY 166 AAGAAATTGACAGATAGAACTGCGCGGTCTATAGTCTCCCTCTCTCTTTGGTGTTCGAA 225  
DB 1507 AAAGGCCACGTGATTCGAACCCGGATGATCCAAATGGAGCCTCAGTTTGGAGGTGCACCC 1566  
QY 226 TGTCACAA 233  
DB 1567 TGCCACGA 1574

RESULT 13  
US-09-976-594-353  
; Sequence 353, Application US/09976594  
; Patent No. 6673549  
; GENERAL INFORMATION:  
; APPLICANT: Furness, Michael  
; APPLICANT: Buchbinder, Jenny  
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS  
; FILE REFERENCE: PA-0041 US  
; CURRENT APPLICATION NUMBER: US/09/976,594  
; CURRENT FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: 60/240,409  
; PRIOR FILING DATE: 2000-10-12  
; NUMBER OF SEQ ID NOS: 1143  
; SOFTWARE: PERL Program  
; SEQ ID NO 353  
; LENGTH: 4866  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. 6673549 415378.3  
US-09-976-594-353

Query Match 0.9%; Score 39.2; DB 4; Length 4866;  
Best Local Similarity 50.5%; Pred. No. 0.25; Indels 0; Gaps 0;  
Matches 95; Conservative 0; Mismatches 93; Indels 0; Gaps 0;  
QY 46 GAATATGCGAACACTTTGCCCTTCAGCCTCTCTACAGAACAGGCTTGCTCTATTCCTTGT 105  
DB 1964 GAACTTGGAGAGTGCATGAGGATCTGGAGCAGGTGGAGAGTGCATGCTCTCTGAATGC 2023  
QY 106 CCCCGGATTTGTAGTATCTGAGTTCTTACCATGCTCCAACTGTAGCAAGGATGTGGG 165  
DB 2024 CCATTTGACTGTGAGTCTACCGAGTGTCCCGAGTGTCCCGAGTGTGGAATGTAAACAGTGCATGTGGG 2083  
QY 166 AAGAAATTGACAGATAGAACTGCGCGGTCTATAGTCTCCCTCTCTCTTTGGTGTTCGAA 225  
DB 2084 AAAGGCCACGTGATTCGAACCCGGATGATCCAAATGGAGCCTCAGTTTGGAGGTGCACCC 2143  
QY 226 TGTCACAA 233  
DB 2144 TGCCACGA 2151

RESULT 14  
US-09-621-976-15639/c  
; Sequence 15639, Application US/09621976  
; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Jobert, S.  
; APPLICANT: Giordano, J.Y.

; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: GENSET.054PR2  
; CURRENT APPLICATION NUMBER: US/09/621,976  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 15639  
; LENGTH: 505  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-621-976-15639  
Query Match 0.8%; Score 37; DB 4; Length 505;  
Best Local Similarity 14.2%; Pred. No. 0.22;  
Matches 55; Conservative 160; Mismatches 170; Indels 2; Gaps 1;  
QY 1089 TCAGGGGAAAAAGGATTTAGAACGAGGAGCGCCATGCTCTCATGGAATCTACAGGGCC 1148  
DB 443 TTAACRWAWAGAGAAWKKWAGGSRACARAGTTWAGACWAMARRGRCARGTGS-- 386  
QY 1149 TGACGGGCAATGCCCTCATTTTGGTGGAGTCTCTTCTGTGAGGATCCAAATGTGCTACCG 1208  
DB 385 WGSWKGGRWGRGMAAAKVRMAAAGGSGYCGMTSYTSGSKMTGKSGMTKTKRMWTTSG 326  
QY 1209 ATGGCTGGCATCAGAGGATCTGTTCCCTCATCATGGAATGTGGCTGGGACATCG 1268  
DB 325 MMWTSYKCTTK 266  
QY 1269 TATTC'GAAAGGCGCTGCGCAAGATGACCGGAGAGATGTATCAGGAGTCTTTTGCC 1328  
DB 265 MMSASAYRARRSMYGARRSMRAGAGWRRARRGKRGKSGMSMSMSMSMSMSMSMSMS 206  
QY 1329 AGTTCCCTCTCTCTGAGAGAGTCTTGTGAATTCCTCCCGAATGAGTGTGTCT 1388  
DB 205 CREMWSCRMYSYSCMGSCRGTC'AKWRYAYAKYASSMGKYMGCRCWCTAKCAR 146  
QY 1389 GAGCAGTGGACGAGTGTCTCTCTGTTCCAGTCTCTTCAATAAACTCAGATGG 1448  
DB 145 MYGYRSRSTGSRGMYRRKMYMYMYMYMYMYMYMYMYMYMYMYMYMYMYMYMYMYMY 86  
QY 1449 GAAACAGACAGTCAAGAACTATCTCT 1475  
DB 85 CKMCSRMMKMS 59

RESULT 15  
US-09-397-992A-25  
; Sequence 25, Application US/09397992A  
; Patent No. 6329175  
; GENERAL INFORMATION:  
; APPLICANT: Conklin, Darrell  
; APPLICANT: Grant, Francis J.  
; APPLICANT: Rixon, Mark W.  
; APPLICANT: Kindvogel, Wayne  
; TITLE OF INVENTION: Interferon-epsilon  
; FILE REFERENCE: 98-46  
; CURRENT APPLICATION NUMBER: US/09/397,992A  
; CURRENT FILING DATE: 1999-09-16  
; PRIOR APPLICATION NUMBER: 60/101,012  
; PRIOR FILING DATE: 1998-09-18  
; PRIOR APPLICATION NUMBER: 60/118,578  
; PRIOR FILING DATE: 1999-02-05  
; PRIOR APPLICATION NUMBER: 60/142,766  
; PRIOR FILING DATE: 1999-07-08  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 25  
; LENGTH: 576  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURES:  
; OTHER INFORMATION: This degenerate sequence encodes the amino acid  
; OTHER INFORMATION: sequence of SEQ ID NO:24.

NAME/KEY: variation  
LOCATION: (1)....(576)  
OTHER INFORMATION: N is any nucleotide.  
IS-09-397-992A-25

Query Match 0.8%; Score 37; DB 4; Length 576;  
Best Local Similarity 30.0%; Pred. No. 0.25;  
Matches 65; Conservative 35; Mismatches 117; Indels 0; Gaps 0;

NY	3988	CGGCGTGCATTCCAGCCTGCAGAAACCTTCTCCTACTGTACACAGGGTGGACTCTGT	4047
DB	233	CNGTGTNCAYGARATHYTNCAFCARATHYTTACNYTNTNCARACNCAVGGNACNATGG	292
NY	4048	GTTGTGAGAAGGGCTATACAGAGATAATGAAATCAAATGGTTTCTCGGATTTACTGCATG	4107
DB	293	GNATHGGGARGARAAYCAVATHGAPARGTNYTNGCNGCNYTNCAYMGCNARYTNGART	352
NY	4108	AAAGTACCAGGCTCAGAGATAAAAGAGCTGATGTGAAAACCTTCTGGGAAAAACAGA	4167
DB	353	AYGTNGARWSNTYNGGNGNYTNAAYGCNGCNCARARWSNGNGNWSNWSNCGNCARA	412
NY	4168	CCTGTCAATTCAAAAATACATGATATTTTAAAGGAT	4204
DB	413	AYTNGNYTNCARATHAAGCNTAYTTYMENGNAT	449

Search completed: February 23, 2004, 20:35:45  
Job time : 336.257 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

M nucleic - nucleic search, using sw model

Run on: February 23, 2004, 17:26:11 ; Search time 1457.49 Seconds  
(without alignments)  
10566.289 Million cell updates/sec

Title: US-10-022-710-1  
Perfect score: 4398  
Sequence: 1 atggcgctgcttcagaa.....atggagacttagacatgtaa 4398

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2308684 seqs, 1750822206 residues

Total number of hits satisfying chosen parameters: 4617368

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.\*

- 1: /cgn2\_6/ptodata/1/pubpna/us07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*
- 16: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*
- 17: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*
- 18: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1232.8	28.0	4810	15	US-10-074-566-40 Sequence 40, Appl
2	1232.8	28.0	5669	15	US-10-074-566-59 Sequence 59, Appl
3	1232.8	28.0	6373	13	US-10-000-512-1 Sequence 1, Appl
4	1232.8	28.0	6373	15	US-10-074-566-1 Sequence 1, Appl
5	1232.8	28.0	6373	15	US-10-074-566-58 Sequence 58, Appl
6	1229.6	28.0	6378	14	US-10-037-270-332 Sequence 332, Appl
7	1229.6	28.0	6378	15	US-10-117-722-332 Sequence 332, Appl
8	580.6	13.2	3053	15	US-10-104-047-958 Sequence 958, Appl
9	512.4	11.7	516	9	US-09-960-253-20 Sequence 20, Appl
10	460	10.5	744	9	US-09-864-761-27172 Sequence 27172, A
11	239.2	5.4	530	14	US-10-029-386-5068 Sequence 5068, Appl
12	235	5.3	235	14	US-10-029-386-18824 Sequence 18824, A
13	201	4.6	583	14	US-10-029-386-8718 Sequence 8718, Appl
14	174	4.0	174	14	US-10-029-386-22453 Sequence 22453, A
15	161.8	3.7	579	9	US-08-919-172-96 Sequence 96, Appl

ALIGNMENTS

RESULT 1

US-10-074-566-40  
; Sequence 40, Application US/10074566  
; Publication No. US20030207348A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimkets, Richard A.  
; APPLICANT: Fernandes, Elma R.  
; APPLICANT: Li, Li  
; APPLICANT: Gorman, Linda  
; APPLICANT: Gusev, Vladimir Y.  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Shenoy, Suresh G.  
; APPLICANT: Spytek, Kimberly A.  
; TITLE OF INVENTION: Polypeptides and Polynucleotides Encoding Same  
; FILE REFERENCE: 15966-556 CIP1  
; CURRENT APPLICATION NUMBER: US/10/074,566  
; CURRENT FILING DATE: 2002-02-13  
; PRIOR APPLICATION NUMBER: 09/619,252  
; PRIOR FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 60/144,722  
; PRIOR FILING DATE: 1999-07-20  
; PRIOR APPLICATION NUMBER: 60/167,785  
; PRIOR FILING DATE: 1999-11-29  
; PRIOR APPLICATION NUMBER: 60/276,994  
; PRIOR FILING DATE: 2001-03-19  
; PRIOR APPLICATION NUMBER: 60/280,898  
; PRIOR FILING DATE: 2001-04-02  
; PRIOR APPLICATION NUMBER: 60/332,241  
; PRIOR FILING DATE: 2001-11-14  
; PRIOR APPLICATION NUMBER: 60/288,062  
; PRIOR FILING DATE: 2001-05-02  
; PRIOR APPLICATION NUMBER: 60/291,766  
; PRIOR FILING DATE: 2001-05-17  
; PRIOR APPLICATION NUMBER: 60/314,007  
; PRIOR FILING DATE: 2001-08-21  
; NUMBER OF SEQ ID NOS: 132  
; SOFTWARE: PatentIn Ver. 2.1

Sequence 22181, Sequence 1739, Ap  
Sequence 15439, A  
Sequence 30452, A  
Sequence 13883, A  
Sequence 6771, Ap  
Sequence 14633, A  
Sequence 85249, A  
Sequence 31224, A  
Sequence 5988, Ap  
Sequence 11407, A  
Sequence 28024, A  
Sequence 22648, A  
Sequence 25110, A  
Sequence 130218, Sequence 402, App  
Sequence 402, App  
Sequence 9643, Ap  
Sequence 249, App  
Sequence 206, App  
Sequence 206, App  
Sequence 206, App  
Sequence 206, App  
Sequence 206, App  
Sequence 184, App  
Sequence 184, App  
Sequence 305, App  
Sequence 184, App  
Sequence 184, App  
Sequence 184, App



2Y 1981 CCACAGAAATGGAGCCCTTGCATCTTAGTGCCAGAGTCTCTCTGCGAGGGAATACCGGC 2040  
2b 2319 ACTCACAATGGCGCAGATGCCAATTAGTCCCTTGGAGCGTGCAACAGACAGCCCTGGA 2378  
2Y 2041 AGCAGTGAACCTGTGGAAAGGGTTACAAACAGAGCTGTCTCATGCACTCTGATGAC 2100  
2b 2379 GCACAGAAAGGCTGTGGCGCTGGCGCAGCAGGCAAGAGCCATTACTTGTGCGCAAGAGAT 2438  
2Y 2101 AACCGGTGACGAGAAATGATGGAATGCTCAAGCAGACAAACGGCATGCCCTCTCTCTTGTG 2160  
2b 2439 GGAGGACAGCTGGAATCCATGAGTGCCTACAGTATGCGAGGCCCTGCGAGCCCTTACC 2498  
2Y 2161 CAGAAATGCAAGTCCCATGTGAGAAAGACTGCACCTTCACTGCTGGTCCAAAGTTTACG 2220  
2b 2499 CAGGCCCTGCCAGATCCCTGCCAGGATGAGTGTCAATTGACACAGCTGGTCCAAAGTTTCT 2558  
2Y 2221 CCCTGCTCCAGCACTGTGAAGCCACAAAAGTAGGGCGGACAGCTTCACAGGGAAGC 2280  
2b 2559 TCATGCAATGAGACTGTGTGAGTTAGGACCAGAAAGCGCACTCTTGTGGAAGAT 2618  
2Y 2281 AGAAAGAGAGAAATCCAGGATTTGACCTTTACCTCTAGTGGAGACAGAACTATGT 2340  
2b 2619 AAAAAGAGGAAATGTAAATATCCCAATTGTATCCCTGATGAGACTCAGTATGT 2678  
2Y 2341 CCTTGTGATGAATTTATATCCACCTTATGGAACTGTGCAATGCTTCTCCAGAA 2400  
2b 2679 CCTTGTGACAAATATATGACAACTGTGGGGAATCTGTGCACTGTATTTTACCAGAG 2738  
2Y 2401 GGCAGAAAGGAGCCCTCACCGAGACTCGGGTACAAGCAGACAGCAAGAAATGTGGAGAA 2460  
2b 2739 GGAAAGTGAAGTGTGCTGGGAATGAAGTACAAGGACACATCAAGGAATCGGACAA 2798  
2Y 2461 GGCCTGGCTTTCCAGCAGATAGCCTTCTGTCTGTATGTAATAAATGGAAGACCTGTGACCCCTCC 2520  
2b 2799 GGATATCGTTACCAAGCAATGGCATGTGTAGCATCAAAATGGCAGGCTGTGGAACATCT 2858  
2Y 2521 TTCTGACGACGCTCTGTTACATTCAGAAATATGTGTCTATCCCTGCCCATTTGATTCG 2580  
2b 2859 AGATGTACAGCCATGTTATCATTTGAGAGGCTGTGATCATCCCTGCCCTCAGACTGC 2918  
2Y 2581 AAGTTAAGCAATCGTCTAGTGTGGGGTCTTGCAGTTTCACTTTGTGGAATGAGATGAGA 2640  
2b 2919 AAGCTCAGTCACTGGTCAACTGGTGGCTGCGCTGACAGCAAGTCTGTGGAGTGGTGTGAAG 2978  
2Y 2641 ATTCGATCCCAATGGCTTAAAGAAACCTTACATGGAGAGGACCATCTGCCCAACTG 2700  
2b 2979 GTTCGTTCTTAATGGCTGCGTGAAGAAACCATATATGGAAGAAAGGCTTGGCCCAACTG 3038  
2Y 2701 GATC-----TCAAGAAATCAGSTACATGAGCAGTCCCATGTTTACAGTGAAGTCAATCAG 2754  
2b 3039 GACCATGTCAACAGGCACAGGTGTATGAGTGTGTCCCATGCCACAGTGAATGCAACCCAG 3098  
2Y 2755 TATTCCTGGTGTAGAACACTGTCTTCATGCAAAATCAACATGAGCTGAGGTGCTG 2814  
2b 3099 TACCTATGGGTACAGAGCCCTGGAGCATCTGCAAGTGAACCTTTGTGAATATGCGGGAG 3158  
2Y 2815 CGCTGTGAGGAGGAAACACAACTPAGAAATCAGATGTGT---GAATCTGCGGATGGT 2871  
2b 3159 AACTGTGGAGAGGCGTGCAACCCGAAAGTGAATGATGCAAGATAAGCAGATGGC 3218  
2Y 2872 GAAGGTGAGCAGTGGATGAGCAACTGTGCAACAGGATGAATTTCCCCAGAAACCCAG 2931  
2b 3219 CCTTCTGAACATGTAGAGGATTAACCTCTGTGACCCAGAAAGATGCCCTGGGCTTAGA 3278  
2Y 2932 TCCTGTTCTTATGTCTCCCAATGAGTGTCTATGCTGAGTGGGACCTTTGAGCAAA 2991  
2b 3279 GTGTGCAAAATACCATGCCCTGAGGACTGTGTGATATCTGAAATGGGGTCCATGAGACCAA 3338  
2Y 2992 TGCCCAAGCATGATGCCATCCCAACAAT---GCAGAGAAAGAACTCGCCACCTGCTAAGA 3048  
2b 3339 TGTGTTTTCCTTCAATCAAGAGCAGTTTCGGGCAAGAGGTGAGTGTATCCCATCAGACAA 3398  
2Y 3049 CCATCACTGAACTCAAGACTTGTGTGAAGACTCAGAGTGGAGCCCTTGCCTCTTGAAT 3108

Db 3399 CCAGTGATGAAGGAAGATCTTGGCCCTAATGCTGTGGAAGAAACCCCTGTAACTGAAC 3458  
QY 3109 GAAATATGCTTCCAGTCCAGTACAATCTAACAGATGGAGCACATGCCAGCTGAGTGA 3168  
Db 3459 AAAAAGTGTCTACCACTATGATTATAATGTAAACAGACTGGAGTACATGTCACTGAGTGA 3518  
QY 3169 AACGCAACCTCTGGTCAAGGCTCAGGACCCGCTGCTTAAGCTGTGTGAGCTGTGTGAGTGGC 3228  
Db 3519 AAGCAGTTTCTGGAAATGGAATTAACCAAGGATGTTGGATTCTGTTGGAAGTATGGC 3578  
QY 3229 AAGCAGTCACTGAGCAATGTGAGCAGCATTAATTTGGAGAAAGCCCGAGAGATGAGC 3288  
Db 3579 AAGTCAGTTGACTGAATATTTGTAAGCGCTTGGCTTGGAGAAAGAACTGGCAGATGAAC 3638  
QY 3289 ATTCCCTCTTGGTGAATGCTCAACTGTGAGCTCTCAGGCTGAGCGCTTGGACA 3348  
Db 3639 AAGTCTGCTGATGGAATGCGCTGTGATGCTGATGCTTCTGATTTGCTCTCTTGGTCA 3698  
QY 3349 GAGTGTTCACAGACTGTGGCCATGGAGGTGGAATGAGCGGACTCGATTTATCATATG 3408  
Db 3699 GAAATGTTCTCAACATGTGGCTCAGAGGAAATGATCCGAGACGAAACAGTGAACCCAG 3758  
QY 3409 CCAACCCAGAGAGAGGACGCGCATGCCCAACAGACTTACCCAGAGAAACCTGCCCA 3468  
Db 3759 CCCTTTCAAGGTGATGGAAGACCATGCCCTTCCCTGATGGACCGATCCAAACCTGCCCA 3818  
QY 3469 GTGACCCCTCTACAGCTGGGTCTTGGCAACTGGTCTGCATGTAAATTTGGAGGTGGA 3528  
Db 3819 GTGAAGCCTTGTATTACGGTGGCAATATGGCAGTGGTCTCCATGCCAAGTGCAGAGGCC 3878  
QY 3529 GACTGTGGGAAGGAGTTTCAGATCGGAGCCTTTCCTGCATGGTCCACAGTGGTTCATA 3588  
Db 3879 CAGTGTGAGAGGAGCAAGCAAGCAATTTCTTGTGTAGTGAAGTGGGTGAGCT 3938  
QY 3589 TCTCATGAGCTGGAGCTGTGAGATGCACTGTGTGGAGAAATGCCCTT----- 3638  
Db 3939 GATGATTTTCAAGCAAGTGTGGAAGAGAAATTTCTGTGCTGACATGAACTCATATAGAT 3998  
QY 3639 -----TCAGGACAGCATCCTGAAGCAGCTGTGTCTGTGCTTCCCTCCAGGAGACTGCCAT 3693  
Db 3999 GGTATAAATAATATGTTTCTGGAGAAATCTCTGAGCCAGCCCTTCCCGAGTGTAT 4058  
QY 3694 TTACAGAGATGTCAGAGTGGAGCAGACATGTGAATTAACCTCATTTGATGGAAGACTTT 3753  
Db 4059 TTGAAGGACTGTCTTCTCTGGAGCCTGTGTGAGTGAATTTGATGAGTGAATCTA 4118  
QY 3754 GAGACTGTGGGCCCGCCAGTCTAGATCAAGGACTTTTAAATTCAGTCTTTTGAAGAACAA 3813  
Db 4119 GGCTTTGGTGAATACAGGTCAAGTCCAGACCGGTGATATAAAGAACTAGAGAAATCAG 4178  
QY 3814 GACAGTCCGCCCCACAGGTTCTAGNAACACGCCCTTGTACAGAGGCAATGTTATCAC 3873  
Db 4179 CATCTGTGCCAGAGCAGATGTTAGAAACAAATCATGTTATGATGAGACGTGCTATGAA 4238  
QY 3874 TACACATGGAAGAGAAAGTCTTTTGGAAACAAATAACGAACTGTATGTCGAGGCTTCA 3933  
Db 4239 TATAAATGATGGCCAGTGTCTTGGAGGGCTCTTCCCGAACAGTGTGTGTCAAAAGTCA 4298  
QY 3934 GATGGCTTAATGTCAAGAGGCTGTCTCCCTCAGGCCCTCTCTGTCCTCATTCGGCAG 3993  
Db 4299 GATGTAATAATGTAAACAGGGGCTGTCTTGTGATGAGCCAGCTGATGCCAGAGTCT 4358  
QY 3994 TGCATTCAGGCTGAGAAACCTTTCTCTCTACTGTACACAGGGTGGAGTCTGTGGTGT 4053  
Db 4359 TGTAAACCCAGTGTAGTCAACCCCACTCTGCTACTGTAGCGAGACAAACATGCCATGT 4418  
QY 4054 GAGAAGGCTATACAGAGATTAATGAATCAAAATGTTTCTTGGATTAATGCA----- 4105  
Db 4419 GAAGAAGGCTACACTGAAGTCAATGTCTTCTTAACAGCACCTTGGACATGACACTTATC 4478  
QY 4106 -----TGAAAGTACAGGCTCAGAGGATAAAGCTGATGTGAAACCTTCTCTG-G 4157



Db 4479 CCCGTGGTGGTATTACCCACCATGGAGGACAAAGAGAGATGTGAAACCAGTCCGGCT 4538  
Qy 4158 GAAAAACAGACCTGTGAATCAAAAAATACATGATATTTTAAAGGA--TGGTCTCTTCAA 4215  
Db 4539 GTACATCAACCCCAACCTCCAGTAACCCAGCAGGAGCGGGAAGGACCTGGTTTCTACAG 4598  
Qy 4216 CCACCTTGATCCAGATGGCGGAGTAAATAATTTGGGTTTATGGCGTTTTCAGGTGGCGCTTTT 4275  
Db 4599 CCAATTGGCCAGATGGGAGACTAAAGACCTGGGTTTACGGTGTAGCAGCTGGGCAATTT 4658  
Qy 4276 CTCATCATGATTTTCTTAATATTTACTTCTTACCTGTGTTCGCAAGAGCCA---AAACCA 4332  
Db 4659 GTGTTACTCATCTTTATTTGTCTCCATGATTTTATCTAGCTTGCATAAAAGCCAAAGAACCC 4718  
Qy 4333 CATCAACACACACCTCCCAACAGAGGCTCTGACCTTTAGCCTTACGATGAGACTTAGAC 4592  
Db 4719 CAAGAAGGCAAAACACCGACTGAACCTTTAACCTTAGCTATGATGAGATGCCGAC 4778  
Qy 4393 ATGTAA 4398  
Db 4779 ATGTAA 4784

RESULT 2  
US-10-074-566-59  
; Sequence 59, Application US/10074566  
; Publication No. US20030207348A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimkets, Richard A.  
; APPLICANT: Fernandes, Elma R.  
; APPLICANT: Li, Li  
; APPLICANT: Gorman, Linda  
; APPLICANT: Gusev, Vladimir Y.  
; APPLICANT: Padigar, Muralidhara  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Shenoy, Suresh G.  
; APPLICANT: Spytek, Kimberly A.  
; TITLE OF INVENTION: Polypeptides and Polynucleotides Encoding Same  
; FILE REFERENCE: 15966-556 CIP1  
; CURRENT APPLICATION NUMBER: US/10/074,566  
; PRIOR FILING DATE: 2002-02-13  
; PRIOR APPLICATION NUMBER: 09/619,252  
; PRIOR FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 60/144,722  
; PRIOR FILING DATE: 1999-07-20  
; PRIOR APPLICATION NUMBER: 60/167,785  
; PRIOR FILING DATE: 1999-11-29  
; PRIOR APPLICATION NUMBER: 60/276,994  
; PRIOR FILING DATE: 2001-03-19  
; PRIOR APPLICATION NUMBER: 60/280,898  
; PRIOR FILING DATE: 2001-04-02  
; PRIOR APPLICATION NUMBER: 60/332,241  
; PRIOR FILING DATE: 2001-11-14  
; PRIOR APPLICATION NUMBER: 60/288,062  
; PRIOR FILING DATE: 2001-05-02  
; PRIOR APPLICATION NUMBER: 60/291,766  
; PRIOR FILING DATE: 2001-05-17  
; PRIOR APPLICATION NUMBER: 60/314,007  
; PRIOR FILING DATE: 2001-08-21  
; NUMBER OF SEQ ID NOS: 132  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 59  
; LENGTH: 5669  
; TYPE: DNA  
; ORGANISM: human  
US-10-074-566-59

Query Match 28.0%; Score 1232.8; DB 15; Length 5669;  
Best Local Similarity 57.7%; Pred. No. 0;  
Matches 2598; Conservative  
10 TGCATTTCAGAGCTGACCGAAGCTGGTTCGAAATGAATATGCAACACTTTCGCTT 69

Db 14 TGCATCCAGAAAGACAAAGACATTCCTCGGAGGATATCATCTGTGAGTACTTTGAGCCC 73  
Qy 70 CAGCCTCTCAGACACAGCTTGGCTCATTCCTGTCCCGGATTTGTAGTATCTGAG 129  
Db 74 AAGCCTCTCCTGGAGCAGCTTGGCTCATTCCTGTCCAGAGAGATTGATCTGTCTGAA 133  
Qy 130 TTCTTACCATGATCCAACTGTAGCAAGGATGTGGGAAGAAATTGCAATAGAACTCGC 189  
Db 134 TTTTCTGCTGCTCGAATGCTTCCAAAGACCTCGCGCAGCGGCTCCAGCACCGGACGGT 193  
Qy 190 GGGGTATAGCTCCCGCTCTCTTGGTGTTCGAAATCTCCAAATCTGACTGAGTCAAGA 249  
Db 194 CATGTGGTGGCGCCCGCAGTTCGGAGGCTCTGGCTGTCCAACTGACGGAGTTCAG 253  
Qy 250 GCGTGTGATGCTCCCATTTCTGTCTCTTGGGGAAGAGATATACATTTAGCCTTAAG 309  
Db 254 GTGTG-----CCAATCCAGTCCATGCGAGGCGGAGGAGCTCAGGTACAGCCTGCAT 304  
Qy 310 GTTGGACCATGGAGTAATGACACTGCTCTATCTTAAAGAAATTAATCCAGCGGAGA 369  
Db 305 GTGGGCGCTGGAGCAGCTGCTCAATGCCCCCTCCGACAGTAAGACAAGGAGGAGA 364  
Qy 370 ACTGTTCTGGAATTTAA-----CTCTGATTTCAAATGAG 402  
Db 365 CCGGGAAGAAATAAGACCGGAAAGGACCGCAGCAAGGAGTAAGGATCCAGAGCC 424  
Qy 403 CGAGTCACCTTTAAACATCAAAAGTTACA-----AAGCACATCATCTTCGAAGTCT 453  
Db 425 CCGAGCTTTATTAAGAAAAGAGAAACAGAAACAGGCAAGACAGAAACAAATAT 484  
Qy 454 TGGGCAATAGATAGGTTATCAAAACCGCGAGGTTTCTGTGTAAGAAGTATGATGACAA 513  
Db 485 TGGGACATCCAGATTGGATATCAGACCAGAGAGTTATGTGATTAACAGACGGGAAA 544  
Qy 514 AATGCTATCTTAAAGCCTTTGCTTCAAGATTCCTTCCCATGACTGTTCAGTCTCTGATC 573  
Db 545 GCTGCTGATTTAAGCTTTTGGCAGCAAGAGAGCTTCCAATGACCTTCCAGTCTGTGTG 604  
Qy 574 ATGCCAAAGACTGTGAACCTCCAGTGGTCTCTCGAGCGCCCTGCTCCAGACATGC 633  
Db 605 ATCACCAGAGTGCAGGTTTCCGAGTGTGTCAGAGTGGGCGCCCTGCTCAAAAACATGC 664  
Qy 634 CGTTCAGGGAGTCTCTTGGCAGGATTTAGGAGCAGGAGCGCGAAGCTGAAGCACATGGT 693  
Db 665 CATGACATGTTGCTCCCTCGAGCACTCGTGTGAAGGACACAGCACTCAGGCACTTCCC 724  
Qy 694 ATTGGAGGTGGAAGAGAGTGTCTGAACTCTTGTGAAGAGCGCTGCAATTGTTGAGGA 753  
Db 725 ATTGGCAGTGAAAAGAGAGTGTCCAGAAATTGGAAGAAAAGAACCCCTGTTGTCTCAGGA 784  
Qy 754 GAACTTCTGCAGCAATGTCCTCAGGATTCCTGAGAACTTCTGAATGGAAGAAATGCCAA 813  
Db 785 GATGGAGTTGTCCCTGTGCCAGTATGGCTGGAGAACTACAGAGTGGACTGAGTCCCT 844  
Qy 814 GTCTCTCTCTCTCCTCCTCAGAGCAGAGAGTCCCACTGTCATGTGACGGGACCCGCTGTGTC 873  
Db 845 GTGGACCCCTTTGCTCAGTTCAGCAGGACAAAGAGCGCGGCAACCCAGACCGCCCTCTCTGGA 904  
Qy 874 GGTGGATCCAGACCGGAGGAGTGTACTGTGCCAG-----AGCGTACCAGCA 921  
Db 905 GGGGCGCATCCAGACCCGAGAGGTGTACTGTGTCAGGCCAAGAAAACCTCTCTCACAA 964  
Qy 922 GCTGCCGCACTGAGGGCCAGAGGAGTCTCTAGACCTGTGTGAAAAGGCAATATGTGTGGGA 981  
Db 965 TTAAGTACCACAAAGAACAAAGAGGCTCAAGCCAAATGAGACTTAAATATATCACTGGA 1024  
Qy 982 CCGCCCGCTTGGCTCTCAGCTCTGCAATATCCCTTGTCTACGGACTGCAATGATCT 1041  
Db 1025 CCTATCCTTAATACTACACAGCTGTGCCACATTCCTTTGCCAACTGAATGTGAATTTCA 1084  
Qy 1042 TCCTGGTTCAGCTGGGCGCTGTGCATCCATGAAAACCTGTGATGAACCTCAGGGGAAAAA 1101  
Db 1085 CTTTGGTTCAGCTGGGCGCTTGTGTTACTTATGAAAACCTGTAATGATCAGCAAGGAAAAA 1144

QY 1102 GGATTTAGAACGAGCGG-----CCATGTCTCATGAAATCTACAGGGCTGCA 1152  
DB 1145 GGCTTCAACTGAGGAAGCGCGCATTAACAATGAGCCCACTGGAGCTCTGGGTAACC 1204  
QY 1153 GGGCAATGGCCCTCATTTGGTGGAGTCTGTTCCTTTGTGAGGATCCAATGTGCTACCCATGG 1212  
DB 1205 GGAACCTGCCCTCATCTACTTGAAGCAATTCCTGTGGAAGAGCTGCCCTGTATGACTGG 1264  
QY 1213 ---CTGCATCAGAAAGGATCTGTTCCCTGTATCATGGA---AATGTGCCCTGGACAT 1266  
DB 1265 AAAGCGGTGAGACTGGGAGACTCGAGCCAGATAACGGAAAGAGTGTGTCTCAGGCACG 1324  
QY 1267 CGTATCTGAAGGCCGTCTCCAGATGACCGCGGAGAGATGTATCAGGGAGTCTTTTC 1326  
DB 1325 CAAGTTCAAGAGGTTGTGTGATCAACAGTGAAGAGAGTGTGACAGACAGCTGTGC 1384  
QY 1327 CCAGTTCCTCCCTCTCTCTGAGAGAGTCTTTGAAATTCCTCGCGAATGAGTGTGTG 1386  
DB 1385 AGAGATGCCATCTTCCCATCCCTGTGGCTGTGATGCCCATAGCCGAAAGACTGTGTG 1444  
QY 1387 CTGAGCGAGTGGAGGAGTGTCTATCTGTTCCTGTCAGTCCCTGTTCAAAATAAAATCAGAT 1446  
DB 1445 CTGAGCATGCTGTAGTGTCTCTGTCTCTGCTACACCTGCTCAGGGAAGACAGAA 1504  
QY 1447 GGGAAACAGACACAGGTCAAGAACTATCTCTGCG---ACTGGCTGGGAAGGTGAAAGCCA 1503  
DB 1505 GGGAAACAGATACAGACAGATCCATCTCTGGCTATGCGGGTGAAGAGGTGAAATTCG 1564  
QY 1504 TGTCCCTCTAGTCAAGTCTCTCAAGACATCTGTTGTTGTAATGACCATCTCTATGCGAG 1563  
DB 1565 TGTCCAATAGCAGTGTCTTGAAGAGTACGAAGCTGTATGAGCATCTTGCACAGTG 1624  
QY 1564 CTTCACCTGGAGACATCGCTTGGGGCCCTGTTCTGAGACACATGTGTAACTGCCCCCT 1623  
DB 1625 TACCCTGGCAACTGTCTCTGGGCCAGTGCAATGGAGCACCTCAGTATCGTCTTTC 1684  
QY 1624 AATGCAACCAATGGCTGGAAATGGAAGACCACTGTGTGTGTAGGCATTCAGACTCGGAGA 1683  
DB 1685 AACACAACTACGATTTGGAATGGGGAGGCCCTCTCTGTCTGGCATGACAGCAAGAAA 1744  
QY 1684 GTCTTCTGTCTCAAGAGTCACTGGGCAAGTAATGACCAAAAGATGTCACAGATTTCTACT 1743  
DB 1745 GTCATCTGTGTGCGAGTCAATGTGGGCCAAGTGGGACCCAAAATGTCCTGAAAGCCCT 1804  
QY 1744 CGACCTGAACCTGTGCGCCCTGTTTCTCCCAATGCAAAAGACTGTATGTGACTGTCT 1803  
DB 1805 CGACCTGAACCTGTGAAGCCCTGTCTGCTTCTTGTGAAGAGAGTGTATTTGTGACCCCA 1864  
QY 1804 TTCACTGAGTGGACACCTCGCCCAAGGATGTCGCAAGCAAGAAATGCCACAGT---AAAA 1860  
DB 1865 TATAGTACTGGACATCATGCCCCCTCTTGTGTGAAGAGGAGTCTCCAGTATCAGGAAG 1924  
QY 1861 CAGTCTCGATACAGAAATCATCATCAAGAAAGCAGCCAAATGAGGCCAGGAATCCAGAT 1920  
DB 1925 CAGTCTAGGCATCGGTCATCATTCAGCTGCGACCAACGGGGGCCGAGACTGCACAGAT 1984  
QY 1921 ACCTTATAGAGAGAGAGTGTGAAGATGTTCTTGTGTCTGTATATCGTGGGAG 1980  
DB 1985 CCCCTCATGAAGAGAGGCGCTGTGGGCACTCAAGCGTGCACAAAGCTACAGGTGGAAG 2044  
QY 1981 CCACAGAAATGGAGCCCTTGATCTTATGTGCGAGAGTCTGTCTGGCAGGGAATAACGGGC 2040  
DB 2045 ACTCACAATGGCCAGATGCCAATTAGTCCCTTGGAGCGTGCACAGACAGCCCTTGA 2104  
QY 2041 AGCAGTGAAGCTGTGAAGAGGGTTTCAAAACAGAGAGTGTCTCATGCACTCTGTATGAC 2100  
DB 2105 GCACAGGAAGGCTGTGGGCCCTTGGCGCAGAGCAAGAGCCATTACTTCTCGCAAGACAGAT 2164  
QY 2101 AACCGGTACAGAGAAATGATGGAATGCTCTAAGCAGACAAAGCAAGCTCTCTCTTGTG 2160  
DB 2165 GGAGCAGAGGTGGAATCCATGAGTGTCTACAGTATGACAGCCCTGTGCGACGCCCTTACC 2224

QY 2161 CAAGATGCAACAGTCCCATGTGCGAAGAGCTGCACCTTCTACTGCTGTGCTCAAGTTTACG 2220  
DB 2225 CAGGCTGCCAGATCCCTCTGCCAGATGACTGTCAATTGACCAGCTGGTCCAAAGTTTCT 2284  
QY 2221 CCCTGCTCCACGAATGTGAAGCCACAAAAGTAGGCGGCGACAGCTCAACAGGGAAGC 2280  
DB 2285 TCATGCAATGGAGACTGTGTGTCAGTTAGGACACAGAAAGCGCACTCTTGTGGAAAAGT 2344  
QY 2281 AGAAGAGGAGAAATGCCAGGATCTTACCTTCTAGTGGAGACAGAACTATGT 2340  
DB 2345 AAAAAGAAAGAAATGTAATAATTCCTTTGTATCCCTGATGAGACTCAGTATGT 2404  
QY 2341 CCTGTGATGAATTTATATCCAACTTATGAAACTGGTCAGATTGCATTTCTTCCAGAA 2400  
DB 2405 CTTTGTGCAAAATATATGCAAACTGTGGGAATGCTGTCAGACTGTATTTTACCAGAG 2464  
QY 2401 GGCAGAGGGAGCCTCACCGAGGACTCGGGTCAACGACAGACAAAGATGTGGAGAA 2460  
DB 2465 GAAAAGTGGAAAGTGTGCTGGGAATGAAAGTACAAGAGACATCAAGGAATGCGACAA 2524  
QY 2461 GGCCTGCGCTTTTCAGAGCAGTAGCTGTCTGATAAAAATGGAAGACCTGTTGACCCCTCC 2520  
DB 2525 GGAATATCGTTACCAAGCAATGCGATGCTACGATCAAAATGSCAGGCTTGTGGAACATCT 2584  
QY 2521 TTCTGCACGACTCTGTTTACATTCAGAAATAATGTGTCTATCCCTGCCCATTTGATGTC 2580  
DB 2585 AGATGTAACAGCCATGTTTACATTTGAGAGGCCCTGCATCTCCCTGCCCTCAGACTGC 2644  
QY 2581 AAGTTAAGCGATTTGTCTAGTTGGGGTCTTGCAGTTTCATCTTGTGAAATTTGGAGTGAGA 2640  
DB 2645 AAGCTCAGTGTGTCACACTGGTTCGCGCTGCAGCAAGTCTGTGGAGTGTGTGAAG 2704  
QY 2641 ATTGCATCCAAATGGCTAAAAACCTTTACAAATGGAGGACGACCATGTGCCAAACTG 2700  
DB 2705 GTTCTGTTCTAAATGGCTCGTGAATAACCATATATATGGAGGAAGGCCCTTGCCTCCAACTG 2764  
QY 2701 GATC-----TCAGAAATCAGTACATGAGGAGTCCCATGTTACAGTGTGCAATCAG 2754  
DB 2765 GACCATGTCAACAGGACAGGTGTATGAGTTGTCCATGCCACAGTGTGCAACAG 2824  
QY 2755 TATTCCTGGTGTAGAACACTGTCTTCATGCAAAATCAACAAATGAGCTGAGGTCCCTG 2814  
DB 2825 TACCCTATGGGTACAGAGCCCTGGAGCATCTGCAAGGTGACCTTTGTGAATATCGGGAG 2884  
QY 2815 CGCTGTGAGAGGACCAACATCTAGGAAATCAGATGTGT---GAATACCTGCCGATGGT 2871  
DB 2885 AACTGTGAGAGGGCGTGCAAACCCGAAAGTGAAGTGCATGCGAAGATACAGCAGATGCG 2944  
QY 2872 GAAGGTGGAGCAGTGGATAGCAACCTGTGCAACAGGATGAAATTCCTCCAGAAACCCAG 2931  
DB 2945 CCTTCTGAACATGTAGAGGATTAACCTCTGTGACCCAGAAAGATGCCCTGGGCTCTAGA 3004  
QY 2932 TCCTGTTCTTATGTGTCCCAATGAGTGTGTCTATGTGTGAGTGGGACCTTTGGAGCAAA 2991  
DB 3005 GTGTGCAAAATACCATGCTGAGGACTGTGTGATATCTGAATGGGTCTCATGACCCAA 3064  
QY 2992 TGCCACACAGTCAATGCGATCCCAACAAAT---GCAGAGAAGAACTCGCCACTGTCTAGA 3048  
DB 3065 TGTGTTTGCCTTGCATCAAAAGCAGTTTCGGCAAGGTGAGTGTATCCCATCAGACAA 3124  
QY 3049 CCATCACTGAATCAAGGACTGTGTGAAAGACTCAGAGTGCAGGCTTGCTCTCTGTAAT 3108  
DB 3125 CCAGCTGATGAAGGAAGATCTTGCCTTAATGTGTGAAAGAAACCTGTAAACCTGAAAC 3184  
QY 3109 GAAATTTCTTCCAGTTCCAGTACAAATTAACAGAGTGGACACATGCGAGTGAAGTCAA 3168  
DB 3185 AAAAATGCTACCTATGATTTAATGTAACAGTGGAGTACATGTCACTGATGTGAG 3244  
QY 3169 AACGCAACCTGTGTCAAGGGCTCAGGACCCGCTGTCTAAGCTGTGTGAGTGTGATGCG 3228  
DB 3245 AAGGAGTTTGTGGAAATGGAATAAAAAACAAGGATGTTGGATTTGTGTCGAAGTGTGCG 3304  
QY 3229 AAGCCAGTCAGCATGGACCAATGTGAGCAGCATTAATTTGGAGAGACCCCAAGAGATGAGC 3288

Db 3305 AAGTCAGTTGACCTGAAATATTTGTAAGCGCTTGGCTGGAGAGAACTGCGAGATGAAC 3364  
Qy 3289 ATTCCCTGCTTGGTGAATGGTGTCACTGCTAGCTCTCAGGTTGAGCGGCTTGGACA 3348  
Db 3365 AGCTCTGATGGTGGATGGCTGTAAGCTGTGAGCTTTCTGATTTGCTTCTCTGTTGCA 3424  
Qy 3349 GAGTGTTCACAGACCTGTGGCCATGAGGTGCAATGAGCCGAGCTCGATTTATCATTTATG 3408  
Db 3425 GAATGTTCTCAACATGTGGCTTCACAGAAATATGATCCGAGACGAAACAGTACCACG 3484  
Qy 3409 CCAACCCAGGAGAGGAGCGGCTGACCCACAGAGCTTACCCAGGAGAAACCTGCCCA 3468  
Db 3485 CCCTTTCAAGGTGATGGAAGACCATGCGCTTCCCTGATGGACCACTGCAAAACCTTGCCCA 3544  
Qy 3469 GTGACCCCTGCTACAGCTGGTCTTGGCAACTGCTGCACTGTGATGTAATTTGGAGGGTGA 3528  
Db 3545 GTGAGACCTTGTATCGGTGGCAATATGCGCAGTGTCTCCATGCCAAGTGCAGGAGCC 3604  
Qy 3529 GACTGTGGGAAGGAGTTCAGATCCGAGCGCTTTCCTGCATGTTGCCACAGTGGTTCAATA 3588  
Db 3605 CAGTGTGGAAGAGGACCAAGAACAGGAACATTTCTTGTAGTAAGTGAATGGTTCAGCT 3664  
Qy 3589 TCTCATGACGTGGAGCTGTGAGGATGCACTGTGTGAGAAATGCGCTT----- 3638  
Db 3665 GATGATTTACGAAAGTGTGTGATGAGAAATCTGTGCTGACATTGAATCATTTATAGAT 3724  
Qy 3639 -----TCAGGACAGCATCTGAAAGCAGCTGTGTTCTGTGCTTCCCGAGGAGCTGCOAT 3693  
Db 3725 GGTAAATAAAATATGTTCTGAGGAATCTGACGACCGCTTGGCCAGGTGACTGTTAT 3784  
Qy 3694 TTAACAGATGCTCAGAGTGGAGCATGTGAATTAACCTGCAATGATGGAAGAGCTTT 3753  
Db 3785 TTGAAGACTGTGTTCTCCGAGCGCTGTGCTGAGTCACTGTGTAATGGTGAAGATCTA 3844  
Qy 3754 GAGACTGTGGCGCCGAGTCTAGATCAAGGACTTTTAAATTCAGTCTTTTGAAGACAA 3813  
Db 3845 GCGTTTGGTGAATACAGCTCAGATCCAGACCGGTGATTAACAGAACTAGAGAAATCAG 3904  
Qy 3814 GACAGTGGCCCGCAACAGTTCTAGAAACACGCGCTTGTACAGGAGGCAATGTTATCAC 3873  
Db 3905 CATCTGTGCCACAGCAGATGTTAGAAACAAATCATGTTATGATGGACAGTGTATGAA 3964  
Qy 3874 TACACATGGAAGCAAGTCTTTGGAAACATAACGAAACGAACTGTATGTTGCCAGGTTCA 3933  
Db 3965 TATAATGATGAGCGAGTCTGTTGGAAGGCTCTTCCGAAAGTGTGTTGTTCAAGGTCA 4024  
Qy 3934 GATGGCTTAAATGTACAGGAGGCTGCTCCCTCAGGCGGCTGCTGCTGCCATTCGGCAG 3993  
Db 4025 GATGTTATAATGTAACAGGGGCTGCTTGGTGTGATGAGCCAGCTGATGCCAGAGTCT 4084  
Qy 3994 TGCAATTCGAGCTGCAGAAACCTTTCTCCTACTGTACACAGGGTGGAGTCTGTGTTGT 4053  
Db 4085 TGTAAACCCAGCTGTAGTCAACCCCACTCGTACTGTGAGGAGACAAACATGCCATTGT 4144  
Qy 4054 GAGAAGGGGTATACAGAGATAATGAATCAAAATGTTTCCCTGGATTTACTGCA----- 4105  
Db 4145 GAAGAAGGGTACACTGAGTCAATGTTCTTAACAGCACCTTTGAGCAATGACACTTATC 4204  
Qy 4106 -----TGAAAGTACAGGCTCAGAGGATTAAGAGCTGTGTAAGAACCTTTCTG-G 4157  
Db 4205 CCGTGTGTGTTATACCCACCTGAGGAGCAAAAGAGGAGATGTGAAACACAGTCCGGCT 4264  
Qy 4158 GAAAAACAGACCTGTGAATTCAAAAATCATGATATTTTAAAGGA--TGGTCTCTTCAA 4215  
Db 4265 GTACATCCAAACCAACCTTCCAGTACCCAGAGGAGCGGGAAGGACCTGTTTCTACAG 4324  
Qy 4216 CCATTTGATCCAGATGCCGAGTAAATAATTTGGTTTATGCGGTTTTCAGGTTGGCGTTT 4275  
Db 4325 CCATTTGGGCGAGATGGAGACTTAAGAACCTTGGTTTACGGTGTAGCAGCTGGGCAATT 4384  
Qy 4276 CTCAATCATGATTTTCTTAATTTACTTCTTCTACTCTGTTTGCAGAGGCA---AAACCA 4332

Db 4385 GTGTACTCATCTTTTATTTGTTCTCCATGATTTATCTAGCTTGCAGAAAGCCAAAGAACCC 4444  
Qy 4333 CATCAAGACACACCTCCCAACAGAAAGCTCTGACCTTAGCTTACGCTAGGAGACTTAGAC 4392  
Db 4445 CAAAGAGGCAAAACAAACCGACTGAAACCTTTAACCTTAGCTATGATGGAGATCCGAC 4504  
Qy 4393 ATGTAA 4398  
Db 4505 ATGTAA 4510

RESULT 3  
US-10-000-512-1  
; Sequence 1, Application US/10000512  
; Publication No. US20020164699A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Bernades, Elma  
; TITLE OF INVENTION: POLYPEPTIDES AND POLYNUCLEOTIDES ENCODING SAME  
; FILE REFERENCE: 15966-556  
; CURRENT APPLICATION NUMBER: US/10/000,512  
; CURRENT FILING DATE: 2001-10-23  
; PRIOR APPLICATION NUMBER: 09/619,252  
; PRIOR FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 60/167,785  
; PRIOR FILING DATE: 1999-11-29  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 6373  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (235)..(4998)  
; NAME/KEY: misc\_feature  
; LOCATION: (6349)  
; OTHER INFORMATION: Wherein n is a or t or g or c.  
US-10-000-512-1

Query Match 28.0%; Score 1232.8; DB 13; Length 6373;  
Best Local Similarity 57.7%; Pred. No. 0; Mismatches 1782; Indels 126; Gaps 17;  
Matches 2598; Conservative 0

Qy 10 TGCATTGAGAGCTGAACCGAAGTGTGGTTGCAATATGAATATGCAACACTTTGCGCTT 69  
Db 505 TGCATCCAGAAAGACAAAGACATTCCTCGGAGGATATCATCTGTGATCTTGGAGCC 564  
Qy 70 CAGCTCTTACAGAACAGGCTTGCCTCATCTTGTCCCGGGATTTGTAGTATCTGAG 129  
Db 565 AAGCTCTCTCGGAGCAGGCTTGCCTCATCTTGTCCAGCAAGATTGCAATCGTGTCTGAA 624  
Qy 130 TTCTTACATGGTCCAACTGTAGCAAGGGATGTGGAGAAATTTGCAGCATAGAACTCGC 189  
Db 625 TTTTCTGCTTGTCCGATGCTCCAGACCTGCGGAGCGGCTCCAGCACCGGAGCGGT 684  
Qy 190 GCGTCTAGCTATCCCTCTCTTTTGTGGTGTGCAATGTCAAATCTGACTGATCAAGA 249  
Db 685 CATGTGTGGCGGCCCGCAGTTCGAGGCTTCGCTGTCTCCAAACCTGACGAGTTCAG 744  
Qy 250 GCCTGTGATGCTCCCATTTCTCTCTTGTGGGAGAGGAAATATACATTTAGCCTTAAG 309  
Db 745 GTGTG-----CCATCCAGTCCATGCGAGCGGAGGAGCTCAGGTACAGCTTGCAT 795  
Qy 310 GTTGGACCATGAGTAAATGCAGACTGCTCATCTTTAAAGAAATTAATCCAAAGCGGAAGA 369  
Db 796 GTGGGCGCTGGAGCACTCTCTCAATGCCCACTCCGCAAGTAAAGACAAGCAAGGAGA 855  
Qy 370 ACTGTTCTGATTTTAA-----CTCTGATTTCAATGAG 402  
Db 856 CGCGGAGGATTAAGAAACCGGAAAGGACCGCAGCAAGAGGATTAAGGATCCAGAGCC 915  
Qy 403 CGAGTCACCTTTTAAACATCAAGTTACA-----AAGCACATCATCATTCGAAAGTCT 453



QY 2581 AAGTTAAGCGATTGGTCTAGTTGGGGCTTTGCGAGTTCACTTTGTGGAATTGGAGTGAGA 2640  
Db 3136 AAGCTCAGTGAAGTGTCAACTGTCGCGCTCGAGCAAGTCCCTGTGGAGTGTGTGAAG 3195  
QY 2641 ATTGATCCAAATGGCTAAAGAAAAACCTTCAATGAGGAGCAACATGTCCTCAACTG 2700  
Db 3196 GTTCGTTCTAAATGGCTCGCGTGAAGAAACCAATATAATGGAGAAAGGCTTCCGCCAACTG 3255  
QY 2701 GATC-----TCAAGAAATCAGGTACATGAGGCGAGTCCCATGTTACAGTGAGTGCATACAG 2754  
Db 3256 GACCATGTCAACACGAGGACAGGTGTATGAGGTTCTCCCATGCCACAGTGAATGCACCCAG 3315  
QY 2755 TATTCCTGGGTTGAGAACACTGCTTTCTATGCAAAATCAACAATGAGTGAAGTCCCTG 2814  
Db 3316 TACCTATGGGTACAGAGCCCTGGAGCATCTGCAAGGTGACCTTTGTGAATATGCGGGAG 3375  
QY 2815 CGCTGTGAGGAGGAACAACAATCTAGGAATAATCAGATGTGT---GAATACTCGGATGGT 2871  
Db 3376 AACTGTGGAGGGGTCGCAACCCGAAAGTGAATGATGCAATGCAAGATGCG 3435  
QY 2872 GAAGTGGAGCAGTGGATGAGCAACCTGTGCAACGAGTGAATTTCCCCAGAAACCCAG 2931  
Db 3436 CCTTCTGAACATGTAGAGGATTAACCTGTGACCCAGAGAGATGCCCCTGGGCTCTAGA 3495  
QY 2932 TCCTGTTCTTATGTGTCTCCCAATGAGTGTGTCTGTCTGAGTGGGACATTTGGAGCAAA 2991  
Db 3496 GTGTGCAAAATTACCATGCGCTGGAGACTGTGTGATATCTGAATGGGTCCATGGACCCAA 3555  
QY 2992 TGCCACACAGTCATGCGATCCCAACACAAAT---GCAGAGAAAGAACTCGCCACCTCTCTAAGA 3048  
Db 3556 TGTGTTTTCCTTGCATCAAAAGAGTTTCCGGCAAGGTGAGCTGATCCCATCAGACAA 3615  
QY 3049 CCATCAGTGAATCAGAGCTTGTGTGAGACTCAAGTGTGAGCTGAGCTTGCCTCTGAAT 3108  
Db 3616 CCAGCTGATGAAGGAAGATCTTGCCCTAATCTGTGAGAAAGAACCTGTAACTGAAAC 3675  
QY 3109 GAAATATGCTTCCAGTCCAGTACAACTTAACAGAGTGGACACATGCGAGCTCAGTGAA 3168  
Db 3676 AAAAATCTCTACCATAATATATGATTAACAGACTGGATGATCATGTCACTGAGTGAG 3735  
QY 3169 AACGCACCTGTGTGCAAGGCGTCAAGACCCGCTGTAACTGTGTGTGAGTGTGATGGC 3228  
Db 3736 AAGGCAGTTGTGGAATATGGAATAAAACAAGAGTGTGGATTGTGTGGAAGTGATGGC 3795  
QY 3229 AAGCCAGTCAGATGAGCCATGTGAGCAGCAGATAATTTGGAGAGACCCAGAGATGAGC 3288  
Db 3796 AAGTCAGTTGACCTGAAATATTTGGAAGCGCTTGGCTTGGAGAGAACTGGCAGATGAAC 3855  
QY 3289 ATTCCCTGCTTGGTGGAAATGCGTGTCAACTGTCAAGCTCTCAGGGTGGACGGCTTGGACA 3348  
Db 3856 ACGTCTGTCATGGTGGAAATGCCCTGTGAACGTGACGCTTTCTGATTGTCTCTCTTGGTCA 3915  
QY 3349 GAGTGTTCAGACACCTTGGCCATGGAGGTGAATGAGCCGAGCTCGATTTATCATTTATG 3408  
Db 3916 GAATGTCTCAAAACATGTGGCTCTCAGGAAATAATGATCCGAAGACGAAACAGTACCCAG 3975  
QY 3409 CCAACCCAGGAGAGGAGCGGCGCATGCCCAACAGAGCTTACCCAGGAGAAACCTTGCCCA 3468  
Db 3976 CCTTTCAAGTGTGAGAGACATGCTCCCTTCCCTGATGGACAGTCCAAACCTTGCCCA 4035  
QY 3469 GTGACCCCTGTCTAGAGTGGTCTTGGCAACTGGTGTGCAATGTAATTTGAGAGGTGGA 3528  
Db 4036 GTCAAGCTTGTATGCGGTGGCAATATGCGCAGTGTCTCCATGTCGAAGTGCAGAGGCC 4095  
QY 3529 GACTGTGGGAGGAGGTTCCAGATCCGACGCTTTCCTGATGGTCCACAGTGGTTCATA 3588  
Db 4096 CAGTGTGGAGAGGAGGACAGAAACAGAAACATTTCTTGTGTAGTGAATGGTGCAGCT 4155  
QY 3589 TCTATGTCAGCTGGAGCTGTCGAGGATGCACTGTGTGGAGAAATGCCCTT----- 3638  
Db 4156 GATGATTTCCAGCAAGTGTGTGATGAGGAATTTCTGTGCTGACATTTGAATCATTTATAGAT 4215

QY 3639 -----TCAGGACAGCATCCTGAAGCAGCTGTGTCTGTGCTTCCCGAGGAGACTGCCAT 3693  
Db 4216 GGTAAATAAAATATGTTCTGTGAGGAATCTTCGACGCCAGCCTTCCCGAGTGTGCTTAT 4275  
QY 3694 TTACAGAGTGTGAGAGTGGAGCAGACATGTGAATTAACCTCATTTGATGGAAGAGCTTT 3753  
Db 4276 TTGAAGGACTGTCTTCTGTGAGCCTGTGTGAGTCACTGTGTGAATGTGTGAGATCTA 4335  
QY 3754 GAGACTGTGGGCGCCAGTCTAGATCAAGGACTTTTATAATTTCAAGTCTTTTGAAGACCA 3813  
Db 4336 GGCCTTGTGGAAATACAGTCCAGATCCAGACCGGTGATTATACAAGAACTAGAGAAATCAG 4395  
QY 3814 GACACTGCCCCCAACAGGTTCTAGAAACACGCTCTGACAGGAGCAATTTGATATCAC 3873  
Db 4396 CATCTGTGCCACAGACAGATGTTAGAAAACAAATCAUATTATGATGGACAGTGTCTATGA 4455  
QY 3874 TACATATGGAAGCAAGTCTTTTGGAAACAAATTAACGAAACGAACTGTATGTCGACGCTTCA 3933  
Db 4456 TATAATGGATGCCAGTGTCTTGGAAAGGCTCTTCCCGAACAGTGTGTGTCAAGAGTCA 4515  
QY 3934 GATGGGTTAATGTCAAGGAGCTGTCTCCCTCAGGCCGCTCTGCTGCTGCTGCCATTCGGCAG 3993  
Db 4516 GATGTTATAAATGTAAACAGGGGCTGTCTTGTGTGATGAGCCAGCCTGATGCCACAGGCT 4575  
QY 3994 TGCATTTCAGCCTGACAGAAACCTTTCTCTCTACTCTGACACAGGGTGGAGTCTGTGTTGT 4053  
Db 4576 TGTAAACCCAGCTGTGTGATCAACCCCTCTGTAAGTGTAGCGAGACAAAACATGCCATTGT 4635  
QY 4054 GAGAAGGCTATACAGAGATTAATGAATCAAAATGATGTTTCTGGATTACTGA----- 4105  
Db 4636 GAAGAAGGCTACACTGAAGTCAATGCTTCTTAACAGCACCTTTGACCAATGCACACTTATC 4695  
QY 4106 -----TGAAAGTACCAGGCTCAGAGATAAAGCTGATGTGAAAACCTTTCTG-G 4157  
Db 4696 CCGTGTGTGTTATCCACCATGAGGACAAAAGAGAGATGTGAAAACCAAGTCGGCT 4755  
QY 4158 GAAAAACAGACCTGTGAATTTCAAAATACATGATATTTTAAAGGA--TGFTCTCTTCAA 4215  
Db 4756 GTATCTCCAAACCAACCTTCCAGTAAACCCAGCAGGACGGGAAGGACCTGGTTTCTACAG 4815  
QY 4216 CCATTTGATCAGATGGCGAGTAAATTTGGGTTTATGGGTTTTCAGTGTGGCTTTT 4275  
Db 4816 CCATTTGGGCGCAGATGGGAGACTAAAGACCTGGGTTTACGGTGTAGCAGCTGGGCAITTT 4875  
QY 4276 CTCATCATGATTTTCTAATATTTACTTCTCTACTTCTTCAAGAGCCA---AAAGCA 4332  
Db 4876 GTGTTACTCATCTTATTTGTCTCCATGATTTATCTAGCTTGCAAAAGCCAAAGAACCC 4935  
QY 4333 CATCAAAGCACACCTCCCAACAGAAAGCTCTGACCTTAGCCTAGATGGAGACTTAGAC 4392  
Db 4936 CAAAGAAGGCAAAACAAACCGACTGAAACCTTTAACCTTAGCCTATGATGGAGATCCGAC 4995  
QY 4393 ATGTAA 4398  
Db 4996 ATGTAA 5001

## RESULT 4

US-10-074-566-1

; Sequence 1, Application US/10074566  
; Publication NO. US20030207348A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimkets, Richard A.  
; APPLICANT: Fernandes, Elma R.  
; APPLICANT: Li, Li  
; APPLICANT: Gorman, Linda  
; APPLICANT: Gusev, Vladimir Y.  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Shenoy, Suresh G.  
; APPLICANT: Spytek, Kimberly A.  
; TITLE OF INVENTION: Polypeptides and Polynucleotides Encoding Same  
; FILE REFERENCES: 15966-556 CIP1

CURRENT APPLICATION NUMBER: US/10/074,566  
CURRENT FILING DATE: 2002-02-13  
PRIOR APPLICATION NUMBER: 09/619,252  
PRIOR FILING DATE: 2000-07-19  
PRIOR APPLICATION NUMBER: 60/144,722  
PRIOR FILING DATE: 1999-07-20  
PRIOR APPLICATION NUMBER: 60/167,785  
PRIOR FILING DATE: 1999-11-29  
PRIOR APPLICATION NUMBER: 60/276,994  
PRIOR FILING DATE: 2001-03-19  
PRIOR APPLICATION NUMBER: 60/280,898  
PRIOR FILING DATE: 2001-04-02  
PRIOR APPLICATION NUMBER: 60/332,241  
PRIOR FILING DATE: 2001-11-14  
PRIOR APPLICATION NUMBER: 60/288,062  
PRIOR FILING DATE: 2001-05-02  
PRIOR APPLICATION NUMBER: 60/291,766  
PRIOR FILING DATE: 2001-05-17  
PRIOR APPLICATION NUMBER: 60/314,007  
PRIOR FILING DATE: 2001-06-21  
NUMBER OF SEQ ID NOS: 132  
SOFTWARE: Patent in Ver. 2.1  
SEQ ID NO 1  
LENGTH: 6373  
TYPE: DNA  
ORGANISM: human  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (6349)  
OTHER INFORMATION: Wherein N is A, or T, or C, or G.  
US-10-074-566-1

Query Match 28.0%; Score 1232.8; DB 15; Length 6373;  
Best Local Similarity 57.7%; Pred. No. 0; Mismatches 1782; Indels 126; Gaps 17;  
Matches 2598; Conservative 0;  
10 TGCATTGAGAGCTGAACCGAGCTGTGGTTCGAAATGAAATATGCGAACACTTTGGCCCTT 69  
505 TGCATCCAGAAAGCAAGACATTCCTCGGAGGATATCATCTGTGAGTACTTTGAGCCC 564  
70 CAGCTCTCTACAGAACAGGCTTGCTCATCTTCCTGTCCTCGGAGTGTGTAGTATCTGAG 129  
565 AAGCTCTCTCGGAGAGGCTTGCTCATCTTCCTGTCGAGAGGATTCATCTGCTGAA 624  
130 TTCTTACCATGCTCCAACTGTAGCAAGGAGTGTGGGAAGAAATTCGAGCATGAACTCGC 189  
625 TTTTCTGCTGCTCGAATGCTCCAGAACTTCGGAGCGGCTCCAGCACCGAGCGGT 684  
190 CGGCTCATAGTCCCTCTCTTTTGGTGTTCGAAATGTCGAAATCTGACTGAGTCAAGA 249  
685 CATGTGTGGCGCCCGGAGCTTCGAGAGGCTCTGGCTGTCCAAACCTGACGGAGTTCAG 744  
250 GCCTGTGATGCTCCCATTTCTGTCCTCTTGGGGAAGAGGAATATACATTTAGCCTTAG 309  
745 GTGTG-----CCAACTCAGTCCATGCGAGCGGAGGCTCAGGTACAGCCTGCA 795  
310 GTTGGACCATGAGTAAATGAGACTGCTCATCTTAAAGAAATTAATCCAAAGCGAAGA 369  
796 GTGGGCGCTGGACACCTGCTCAATGCCCACTCCCGACAAGTAAGACAAGCAAGGAGA 855  
370 ACTGTTCTGATTTTAA-----CTCTGATTTCAATGAG 402  
856 CGCGGGAAGATAAAGACCGGAAGAGACCGGAGCAAGAGGTAAGAGGATCCAGAGCC 915  
403 CGAGTCCACTTTAAACATCAAGGTTACA-----AAGCACATCATCTTGAAGTCT 453  
916 CGCAGCTTATTAAGAAAGAGAGAAACAGAAACAGGCAAGACAGACAGAGACAATAT 975  
454 TGGCAATAGAGATGAGTTATCAACCCGAGGTTTCTGTGTAAGAAAGTATGACAA 513  
976 TGGGACATCCAGATTGATATCAGCAGAGAGGTTATGTGCAATTAACAAGACGGGAAA 1035  
514 AATGCTATGTTAAGCCCTTGCCTTCAGATGCTTCTTCCCATGACTGTTCACTCTGCATC 573

Db 1036 GCTGCTGATTTAAGCTTTTGGCAGCAAGAGAAGCTTCCAAATGACCTTCCAGTCTGTGTG 1095  
QY 574 ATGCCCAAGAGCTGTGAACCTTCCAGTGGTCTCTCTGGAGCCCTGTCTCAAGACATGC 633  
Db 1096 ATCACAAAGAGTGCACAGTTTCCAGTGGTTCAGAGTGGAGCCCTGTCTCAAAACATGC 1155  
QY 634 CGTTCAGGAGTCTCTTCCAGGATTTAGGAGCAGGAGCCGGAACGTAAGACATGCT 693  
Db 1156 CATGACATGTTGCTCCCTCGCAGCACTCGTGTGAAGACACGAACCATCAGGCAGTTTCCC 1215  
QY 694 ATTGAGGTGAAGAGTGTCTGAACCTTCTGAGAAAGAGGCTGCATTTGTGAAGA 753  
Db 1216 ATTGAGGTGAAGAGTGTCTGAACCTTCTGAGAAAGAGGCTGCATTTGTCTCAAGA 1275  
QY 754 GAACTTCTGACAGCAATGTCAGGATTTCTGAGAACTTCTGAAATGGAAGAATGCCAA 813  
Db 1276 GATGAGTGTCTCCCTGTGCCAGCTATGCTGGAGAACTACAGAGTGCAGTGCCT 1335  
QY 814 GTCTCTCTCTCTGAGCAGAGGATCCCACTGCGCATGTGACGGGACCCGTGTGTGCG 873  
Db 1336 GTGGACCTTTGCTCAGTCAGCAGGACAAAGAGCGCGCAACCAAGACGCGCTCTGTGGA 1395  
QY 874 GGTGGATCCAGACCCGAGGTGTACTGTGCCAG-----AGGTACACGA 921  
Db 1396 GGGGCACTCCAGACCCGAGAGGTGTACTGCTGAGGCCCAAGAAACCTCTCTCAAA 1455  
QY 922 GCTGCCGACTGAGGCGCAAGAAAGTCTTAGACCTGTGGAAGAGGCAATATGTGTGGA 981  
Db 1456 TTAAGTAGCCCAAGAAACAAAGAGGCTTCAAGGCAATGGACTTAAATATATGCACTGA 1515  
QY 982 CCGGCGCTGCTCTCAGCTGTCAATATCCCTGCTCTAGGAGCTGATGATGATGATGATG 1041  
Db 1516 CTATCTCTCTCTCTGAGCAGCTGTGCAATCTCTCTGCACTGAAATGAAAGTTTCA 1575  
QY 1042 TCCTGCTCAGCTGCGGCTGTGCAATCCATGAAACTGTCTATGAACTCTCAGGGGAAAAA 1101  
Db 1576 CTTGCTCAGCTTGGGACCTTGTACTTATGAAACTGTATGATATCAGAAAGGAAAAA 1635  
QY 1102 GGAATTAGAACGAGCAGCG-----CATGCTCTCATGGAATCTACAGGCGCTGA 1152  
Db 1636 GGCCTTCAAACTGAGGAGCGGCGCATTTACCAATGAGCCACTGGAGGCTCTGGGGTAACC 1695  
QY 1153 GGGCATTCGCTCATTTGGTGGAGTCTCTCTTGTGAGGATCCAAATGCTGCTACCGATGG 1212  
Db 1696 GGAACATTCGCTCTCTCTGAGAGGCTCTCTCTGAGAGGCTGCTCTGATGACTGG 1755  
QY 1213 ---CTGGCATCAGAAAGGATCTGTTTCTCTGATCATGAA---AATGTCCTGGGACAT 1266  
Db 1756 AAAGCGGTGAGACTGGGAGACTGGGAGCCAGATAACCGAAAGAGTGTGTCCAGGACG 1815  
QY 1267 CGTATTTGAAAGCGCTCTGCCAGAAATGACCGGAGAGATGTATCAGGAGTCTTTGC 1326  
Db 1816 CAAHTTCAAGAGTGTGTGTCATCAAAGTATGAGAAAGTGTGACAGACAGTGTGC 1875  
QY 1327 CCAGTTCCCTCTCTCTGAGAGGAGTCTTGTGAATTTCCCTGCGGAATGGAGTGTGTG 1386  
Db 1876 AGAGATGCCATCTTCCCATCCCTGTGGCTGTGATGCCCATGCCGAAAGACTGTGTG 1935  
QY 1387 CTGAGCAGTGAAGAGTGTGTGTCATCTCTGTTCCAGTCTCTGTTCAATATAAACTCAGAT 1446  
Db 1936 CTCACACATGTTTACGTTGCTCTCTGCTTCTGCTTCTGCTTCTGCTGAGGAAACGACAGAA 1995  
QY 1447 GGGAAACAGACCCAGCTCAAGAACTATCTGGC---ACTGGCTGGGAGAGTGTGAAGCA 1503  
Db 1996 GGGAAACAGATACAGACAGATCCATTTCTGCTCTATGCGGTGAGAGGTGGAATTCGC 2055  
QY 1504 TGTCCCTCTAGTCTGAGGCTCTCTCAAGAGCATCGTTTGTGTAATGACCATCTCTGTATGAG 1563  
Db 2056 TGTCCAAATAGCAGTGTCTTTCAGAAAGTACGAAAGCTGTATGAGCATCTCTTGCACAGTG 2115  
QY 1564 CTTTCACTGGGAGACATCGCTTGGGCGCTTGTCTGAGGACACATTTGTAACCTGCTT 1623



Db 2116 TACCAGTGGCAAACTGGTCCCTGGGGCCAGTGCATGTGAGGACACCTCAGTATCGTCCTTC 2175  
 Qy 1624 AATGCAACCAATGGCTGGAAATGGAGAGCCAGTGTGGTGTAGGCATTCAGACTCGAGA 1683  
 Db 2176 AACACAACTACGACTTGGAAATGGGAGGCCCTCTCTCTGTGCGCATGCACAAAGAAA 2235  
 Qy 1684 GTCTTCTGTGTCAAGAGTCACTGGGACAGTAATGACCAAAAGATGTCAGATCTTACT 1743  
 Db 2236 GTCACTGTGTGCGAGTCAATGTGGGCCCAAGTGGGACCCCAAAAATGTCTCTGAAAGCCTT 2295  
 Qy 1744 CGACCTGAAACTGTGCGGCCCTGTCTTCTCCCATGCAAAAAGACTGTATTGTGACTGCT 1803  
 Db 2296 CGACCTGAAACTGTAAAGCCTGTCTCTCTCTGTAAGAGGACTGTATTGTGACCCCA 2355  
 Qy 1804 TTCAGTGTAGGACACCTGCGCCCAAGAGTGTGCCAAGCAGGAAATGCCACAGT---AAAA 1860  
 Db 2356 TATAGTGTAGTGCATCATGCTCCCTCTTCTGTGTAAGAGGAGGACTCCAGTATCAGGAAG 2415  
 Qy 1861 CAGTCTCGATCAGAAATCATCATCAAGAGCAGCAATGAGGCGCAGGAATGCCAGAT 1920  
 Db 2416 CAGTCTAGGATCGGGTCAATCAITCAGTGCAGCCCAAGCGGGGCCGAGACTGCACAGAT 2475  
 Qy 1921 ACCTTATATGAGGAGAGAGTGTGAAGATGTTCTTGTCTCTGTATATCGGTGGAAG 1980  
 Db 2476 CCCCTCTATGAAGAGAGAGGCTGTGAGGCACTCAAGCGTGCCTAAAGCTTACAGGTGGAAG 2535  
 Qy 1981 CCACGAATGAGCCCTTGCATCTTAGTGCCAGAGTCTGTCTGCGCAGGGAATACCGGC 2040  
 Db 2536 ACTCAAAATGGCGAGATGCCAAATAGTCCCTGGAGCGTGCACAGAGAGCGCCCTGGA 2595  
 Qy 2041 AGCAGTGAAGCCTGTGGAAAGGGGTACAAAACAGAGTGTCTCATGCTATCTCTGATGAC 2100  
 Db 2596 GCACAGGAGGCTGTGGGCTGGCGCAGGCAAGGCCATTAATCTGTGCGAAGCAGAT 2655  
 Qy 2101 AACCGGTACGAGAAATGTAATGCAATGCTCAAGAGAGACAAAGCGCATGCCCTCTCTGTTG 2160  
 Db 2656 GGAGGACAGGCTGGAATCCATGAGTGCCTACAGTATGAGGCGCTGTGCCAGCCCTTACC 2715  
 Qy 2161 CAAGAAATGACAGTCCCATGTGCGAAGAGTGCACCTTCACTGTGTGCTCAAGTTTACG 2220  
 Db 2716 CAGGCTGCCAGATCCCTGCGCAGAGTACTGTCAATGACAGCTGTGCTCAAGTTTCT 2775  
 Qy 2221 CCTGTCTCAGAACTGTGAAGCCACAAAGATGAGCGCGCAGAGCTCACAGGGAAGAC 2280  
 Db 2776 TCATGCAATGAGAGAGTGTGGTGCAGTTAGGACCAAGAGCGCACTCTTGTGGAAGAAAGT 2835  
 Qy 2281 AGAAGAGGAGAAATGCCAGATCTGACCTTTACCTTACCTGTAGTGAGAGAGACTATGT 2340  
 Db 2836 AAAAGAGGAGAAATGTAAGAAATTTCCCATTTGTATCCCTGANTGAGACTCAGTATTGT 2895  
 Qy 2341 CCTGTGTGAATTTATATCCCAACCTTATGGAAGTGTGTGAGATTTGCAATCTTCCAGAA 2400  
 Db 2896 CTTGTGACAAATATAATGACAACTGTGGGGAATGTGTGAGACTGTATTATTACAGAG 2955  
 Qy 2401 GGCAGAGGAGGACCTCACGAGACTGGGGGTACAGCAGACAGCAAGAAATGTGGAGAA 2460  
 Db 2956 GGAAGAGTGGAAAGTGTGTGGGAATGAAAGTACAGGAGACATCAAGGAATGCGGACAA 3015  
 Qy 2461 GGCCTGCGCTTTCGAGCAGTAGCCTGTCTGATAAAAATGGAAGACCTGTGACCCCTCC 2520  
 Db 3016 GGATATCGTTACCAAGCAATGGATGTACGATCAAAATGCGAGGCTGTGTGGAACATCT 3075  
 Qy 2521 TTCTGAGCAGCTGTGTGTACATTCAGAAATATGTGTCAATTCCTGCGCCATTTGATTCG 2580  
 Db 3076 AGATGTAAACAGCCATGTTTACATGAGGAGGCTGCATCATCCCTGCGCCCTCAGACTGC 3135  
 Qy 2581 AAGTTAAGGATTTGTGTAGTTGGGGCTTCTGAGTTCATCTTGTGGAATGGAGTGA 2640  
 Db 3136 AAGCTCAGTGTAGTGTCCAACTGTGTGCGCTGCACAAAGTCTCTGTGGAGTGGTGTGAAG 3195  
 Qy 2641 ATTTCGATCCAAATGGCTAAAGAAACCTTACAAATGGAGGACGACCATGTCCCAAACTG 2700  
 Db 3196 GTTGTGTTCTAAATGGCTGTGGAAGAAACCATATATATGAGGAGGAGGCTTGGCCCCAACTG 3255

Qy 2701 GATC-----TCAAGAAATCAGGTATCATGAGGAGTCCCATTTACAGTGTGCAATCAG 2754  
 Db 3256 GACCATGTCAACAGGACACAGGTGTATGAGGTGTGCCATGCCACAGTGTGCAACAG 3315  
 Qy 2755 TATTCCTGGGTGTAGAACACATGCTTCTCATGCAAAATCAACAAATGAGTGTAGGTCCTG 2814  
 Db 3316 TACTTATGGGTACAGAGCCCTGGAGCATCTCAAGAGTGACCTTTGTGAATATGCGGGAG 3375  
 Qy 2815 CGCTGTGGAGGAGAACACAAATCTAGAAAATCAGATGTGT---GAATCTCGGATGCT 2871  
 Db 3376 AACTGTGGAGGGCTGCAACCCGAAAGTGAGATGCATGCAGAAATACAGCAGATGAC 3435  
 Qy 2872 GAAGGTGGAGAGTGTGATAGCAACCTGTGCAACAGAGATGAAATTTCCCCAGAAACCCAG 2931  
 Db 3436 CCTTCTGAAACATGTAGAGGATTAACCTCTGTGACCCAGAAAGATGCCCTGGGCTTAGA 3495  
 Qy 2932 TCCTGTCTCTTATGTGTGCCAATGATGTGTCTGTGAGTGGGACTTTTGGAGCAAA 2991  
 Db 3496 GTGTGCAAAATACCATGCCCTGAGGACTGTGTGATATCTGAAATGGGTCCATGGACCCAA 3555  
 Qy 2992 TGCCCAAGTCTATGCGATCCCCACACAAT---GCAGAGAAAGAACTGCGCACTGTCTAAGA 3048  
 Db 3556 TGTGTTTTCCTTGCATCAAAAGCAGTTTCCGCGCAAGGTGAGTGCATGCCATCAGACAA 3615  
 Qy 3049 CCATCACTGAACTCAAGGACTTGTGCTGAAGACTCACAGGTGCGCCTTGCCTCTGAT 3108  
 Db 3616 CCAGCTGTAGGAAGATCTTGGCCCTAATGCTGTGTGAAAGAAACCTTGAACCTGAAC 3675  
 Qy 3109 GAAAAATGGCTTCCAGTTCAGATCAATCTAAACAGAGTGGAGCACATGCCAGCTGAGTGAA 3168  
 Db 3676 AAAAACTGTCTACCATATGATTAATGTAACAGACTGGAGTACATGTCACTGAGTGAG 3735  
 Qy 3169 AACGCACTGTGTGTCAGGCGTCAAGGCGTCAAGGCGCTGCTTAAGCTGTGTGTCAGTGTGCG 3228  
 Db 3736 AAGCGAGTTTGTGGAATGGAATAAAAACAAGATGTTGGAATTTGTTTCAAGAGTGTGCG 3795  
 Qy 3229 AAGCGAGTCAAGATGCAACCAATGTGAGCAGCAATTAATTTGGAGAGCCGCCAGAGAAATGAGC 3288  
 Db 3796 AAGTCAGTTGACCTGAAATATTTGTAAGCGCTTGGCTTGGAGAAAGAACTGGCAGATGAAC 3855  
 Qy 3289 ATTCCCTGTGTGTGAATGCGTGTGTCATCTGTGAGCTCTCAGGCTGAGCGGCTTGGACA 3348  
 Db 3856 ACCTCTGCTGATGGTGAATGCCCTGTGAACCTGTGAGCTTCTGATTTGCTCTCTGTTGCTCA 3915  
 Qy 3349 GAGTGTTCACAGACCTGTGCGCATGAGAGTTCGAATGAGCGGCACTCGATTTATCATATG 3408  
 Db 3916 GAATGTTCTCAACATGTGGCCCTCACAGGAATAATGATCCGAAAGAGCAACAGTGTGCCAG 3975  
 Qy 3409 CCAACCCAGGAGAGGACCGGCATGCCCCACAGAGCTTTACCCAGAGAGAAACCTTGCCCA 3468  
 Db 3976 CCCTTTCAAGGTGATGGAAGACCATGCCCTTCCCTGATGGAACAGTCCCAACCTTGCCCA 4035  
 Qy 3469 GTACCCCTCTGTACAGCTGGGTCTTGGCAACTGGTCTGCACTGTGATTAATTTGGAGGTGA 3528  
 Db 4036 GTGAAGCCTTGTATCGTGGCAATATGSCCAGTGTCTCCATGCGCAAGTGTGAGGAGGCC 4095  
 Qy 3529 GACTGTGGGGAAGGAGTTCAGATCCGACGCTTTTCTGCACTGCTCCACAGTGTGTTCAATA 3588  
 Db 4096 CAGTGTGGAGAGGGACCAAGAACATTTCTTGTGTAGTAAGTGTGAGTGTGAGTGTGAGTGT 4155  
 Qy 3589 TCTCATGTGAGTGTGAGTGTGAGGATGCACTGTGTGAGGAAATGCGCCTT----- 3638  
 Db 4156 GATGATTTTCAAGAAAGTGTGTGATGAGGAAATTTCTGTGCTGACATTTGAATTCATTATAGAT 4215  
 Qy 3639 -----TCAGGACAGCATCTGAGGAGTGTGTCTGTGCTTGGCCAGGAGACTGCCAT 3693  
 Db 4216 GGTAAATAAATAATGTTCTGGAGGATCCTGAGGAGCCCTTGGCCAGGCTGACTGTAT 4275  
 Qy 3694 TTAAACAGAAATGTGAGTGTGAGGACATGTGAATTAACCTGCAATGTGAGGAGAAAGCTTT 3753  
 Db 4276 TTGAAGAGTGTGCTTCTCGAGGCTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 4335

3754	GA	AG	CTGTGGCGCCGAGTCTAGATCAAGGACTTTTATTAATTCAGTCTTTTGAGAACCAA	38113
4336	GG	CTTTGGTGGAAATACAGGTCCAGTCCAGACCGGTGATTATATCAAGAACTAGAGAATCAG	43955	
3814	GAC	AGTGTCCCCCAACACAGGTCTTAGAAACACGGCCCTGTGTACAGGAGGCAAAATGTTATCAC	3873	
4396	CAT	CTGTGCCAGACGAGTGTAGAAACAAATCATGTTATGATGCGACAGTGCTATGAA	4455	
3874	TAC	CATGGAAGACAGTCTTTTGGAAACAATAACGAACGAATGTTATGGTGCCAGCGTTCA	3933	
4456	TATA	ATGATGATGCCAGTGTCTTGGAAAGGGCTCTTCCGGAACAGTGTGTGTCAAAGGTCA	4515	
3934	GAT	GGGTTAAATGTACACAGGAGCTGCTCCCTCAGGCGCGTCTGCTGCCATTCGGCAG	3993	
4516	GAT	GTATATAAATGTAAACAGGGGCTCTGTGTGATGAGCAGGCTGATCCGACAGGTCT	4575	
3994	TGC	ATTCACGCTGCAGAAAACTTTTCTCTACTGTACACAGGGTGGAGTCTGTGGTTGT	4053	
4576	TGTA	ACCCACCGTGTAGTCAACCCCACTCGTACTGTAGCGAGACAAAAACATGCCATTGT	4635	
4054	GAGA	GGGCTATACAGAGATAATGAATCAATGGTTTCTCTGGATTACTTGCA-----	4105	
4636	GAGA	AGGGGTACACTGAAGTCATGCTCTTAAACAGCACCCCTTGAGCAATGCACACTTATC	4695	
4106	-----	TGAAAGTACACGGCTCAGAGGTAAAAAGCTGATGTGAAAAACCTTTCTG-G	4157	
4696	CCG	TGGTGGTATTATCCACCATGGAGGACAAAGAGGAGATGTGAAAACCACTGGGCT	4755	
4158	GAAA	ACAGACCTGTGAATTCAAAAATACATGATATTTTTTAAAGGA--TGSTCTCTTCAA	4215	
4756	GTAC	TCCAAACCCCTCCAGTAAACCCAGCAGGACGGGGAAGGACCTGGTTCTTACAG	4815	
4216	CCACT	TGATCCAGATGGCGAGTAAAAATTTGGGTTTATGGCGTTTTCAGGTGGCGTTT	4275	
4816	CCATT	GGGCCAGATGGGAGACTTAAGACCTGGGTTTACGTTGTAGCAGCTGGGGCATTT	4875	
4276	CTC	ATCATGATTTTCTTAATTTATCTTCTACTGTTTGGTTCGAAGAAGCCA---AAACCA	4332	
4876	GTGT	ACTCATCTTTTATTTGTCTCCATGATTTATCTAGCTTGCAAAAAAGCCAAAGAAACCC	4935	
4333	CAT	CAAGCACACCTCCCAACAGAGACCTCTACACCTTAGCCCTACGATGGAGACTTAGAC	4392	
4936	CAAA	GAAGGCAAAACACCGACTGAAACCTTTAACTTTAGCCCTATGATGAGATGCGGAC	4995	
4393	ATG	TAA 4398		
4996	ATG	TAA 5001		

/	PRIOR APPLICATION NUMBER: 60/276,994	
/	PRIOR FILING DATE: 2001-03-19	
/	PRIOR APPLICATION NUMBER: 60/280,898	
/	PRIOR FILING DATE: 2001-04-02	
/	PRIOR APPLICATION NUMBER: 60/332,241	
/	PRIOR FILING DATE: 2001-11-14	
/	PRIOR APPLICATION NUMBER: 60/288,062	
/	PRIOR FILING DATE: 2001-05-02	
/	PRIOR APPLICATION NUMBER: 60/291,766	
/	PRIOR FILING DATE: 2001-05-17	
/	PRIOR APPLICATION NUMBER: 60/314,007	
/	PRIOR FILING DATE: 2001-08-21	
/	NUMBER OF SEQ ID NOS: 132	
/	SOFTWARE: PatentIn Ver. 2.1	
/	SEQ ID NO 58	
/	LENGTH: 6373	
/	TYPE: DNA	
/	ORGANISM: human	
/	FEATURE:	
/	NAME/KEY: misc feature	
/	LOCATION: (6349)	
/	OTHER INFORMATION: Wherein N is A, or T, or C, or G.	
/	US-10-074-566-58	

  

Query Match	28.0%;	Score 1232.8;	DB 15;	Length 6373;
Best Local Similarity	57.7%;	Pred. No. 0;		
Matches 2598; Conservative	0;	Mismatches 1782;	Indels 126;	Gaps 17;

  

Qy	10	TGCAATTCAGAGCTGAACCGAACTGTGGTTGCCAATGAATAATGCAGAACACTTTGGCCCTT	69
Dd	505	TGCATCCAGAAAGACAAGAATCATCTCGGAGGATATCATCTGTGAGTACTTTTGAGCCC	564
Qy	70	CAGCCTCTACAGAACAGAGCTTGCCTCATCTCTGTGCCCGGATTGTGTAGTATCTGAG	129
Dd	565	AAGCCTCTCTGGAGCARGGTTCCTCATCTCTGTGCCAGAGATTGCATCGTGTCTGAA	624
Qy	130	TTCTTACCATGFTCCAACTGTAGCAAGGATGTGGAGAGAAATTTGAGCATAGAACTCGC	189
Dd	625	TTTTCTGCCTGTGTCCGAATGCTCCAAGACTGTGGCAGCGGGCTCCAGCACCGGAGCGCT	684
Qy	190	CGCGTCATAGCTCCCCCTCTCTTTGGTGGTTTGCATGTCCAAATCTGACTGAGTCAAGA	249
Dd	685	CATGTGGTGGCGCCCGCCGAGTTTGGAGGCTGTGGCTGTCCAACTGACGGAGTTCCAG	744
Qy	250	GCCTGTGATGTCCCATTTCTCTCTCTTGGGGAAGAGGAATATACATTTAGCCTTAAG	309
Dd	745	GTGTG-----CCAATCCAGTCCATCGAGGCGGAGGCTCAGGTACAGCCTGCAT	795
Qy	310	GTGGACCATGGAGTAAATGCAGACTGCCTCATCTTTAAGAAATTAATCCAAGCGGAAGA	359
Dd	796	GTGGGGCCCTTGGAGCACCTGCTCAATGCCCCACTCCGACAAAGTAAGACAAGCAAGAGA	855
Qy	370	ACTGTCTGTGATTTTAA-----CTCTGATTTCAAATGAG	402
Dd	856	CGCGGNAGATTAAGACGGGAAAAGGACCGCAGCAAGGAGTAAGGATCCAGAAGCC	915
Qy	403	CGAGTCACCTTTAAACATCAAGTTACA-----AAGCATCATCATTCGAAGTCT	453
Dd	916	CGCGAGCTTATTAAAGAAAAGAGAAAAACAGAAACAGGCAGACAGACAAGAGAACAAATAT	975
Qy	454	TGGGCAATAGAGATAGTTATCAAAACCGGACGGTTTCGTGTACAAGAGTGTATGCACAA	513
Dd	976	TGGACATCCAGATTGGATATCAGACGAGAGTTTATGTGCATTAACAGACGGGAAA	1035
Qy	514	AATGCTATGTTAAGCCTTTGCCCTTCAAGATTCTCTCCCATTTGACTGTTTAGTCTCTGATC	573
Dd	1036	GCTGCTGATTTAAGCTTTTGGCCAGCAAGAGAACTTCCAATGACCTTCCAGTCTCTGTGTG	1095
Qy	574	ATGCCCAAGACTGTGAACCTCCAGTGTGTCTCTCTGGAGCCCTGCTCCAGACATGC	633
Dd	1096	ATCACCAGAGGTGCCAGGTTTCCGAGTGTGTAGATGGAGCCCTCTGCTCAAAAACATGC	1155
Qy	634	CGTTCAGGGAGTCTCTTGCCAGGATTTAGAGCAGGAGCCGGAACGTGAAGCACATGGCT	693

1156 CATGACATGTTGTCCTCCCTGCGAGCACTCGTGTAAAGACACGACCAATCAGCGAGTTTCCC 1215  
694 ATTGGAGTGGAAAGAGTGTCTCTGAACTTCTTGAGAAAGAGCGCTGCATTTGTTGAGGA 753  
1216 ATTGGCAGTGAAGAGAGTGTGAGAAATTTGAAGAAAGAAACCTGTTTGTCTCAAGA 1275  
754 GAACTTCTGAGCAATGTCACAGATATCTCTGAGAACTTCTGAATGGAAAGAAATGCCAA 813  
1276 GATGGAGTGTGTCCTCTGCGACCATGATGCTGAGAACTACAGATGGACTGAGTSCCGT 1335  
814 GTCTCTCTCTCTGAGAGCAGAGATCCCACTGGCATGTGACGGGACCCGTGTGTGGC 873  
1336 GTGACACCTTTGCTCAGTCAGCAGGACAAAGAGCGGGCAACACAGACGGCCCTCTGTGA 1395  
874 GGTGGGATCCAGACCCGGGAGGTGTACTGTGCGCAG-----AGCGTACCAGCA 921  
1396 GGGGGCATCCAGACCCGAGAGGTGTACTGTGCGAGGCCAACGAAACCTCTCTCACAA 1455  
922 GTCGCCGCTAGGGGCAAGAGTCTCTAGACCTGTGGAAGGCAATATGTGGGA 981  
1456 TTAAGTACCCACAGAAACAAAGAGCCCTCAAGCCCAATGACTTAAATTTATGCACTGA 1515  
982 CCCGCCGTTGCCCTCTCAGCTCTGCAATATCCCTTCTCTACGAGCTGCAATGATCT 1041  
1516 CCTATCCCTAATCTACAGAGCTGTGCCATTTCTTGTCTCACTGATGTGAAGTTTCA 1575  
1042 TCTGTGTGAGCTGGGCTGTGATCCATCAAACTGTATCAAACTCAGGGGAAAAA 1101  
1576 CTTTGGTCACTGGGACCTGTACTTATGAAAACTGTAATGATCAGCAAGGAAAAA 1635  
1102 GGATTTAGACAGGCGAG-----CCATGCTCTCATGTGAATCTACAGGCGCTGCA 1152  
1636 GGCTTCAAACTGAGGAAGCGCGCAATTAACAAATGAGCCCACTGGAGCTCTGGGGTAA 1695  
1153 GGGCAATGCTCATTTGGTGGAGTCTGTCTCTGTGAGGATCCAAATGCTCTACCGATGG 1212  
1696 GGAATCTGCTCACTTACTTGGAGCCATCTCCCTGTGAAGAGCTGCTGTTATGACTGG 1755  
1213 ---CTGGATCAGAAAGATCTGTTCCCTGTATCATGGAA---AATGTGCTGGGACAT 1266  
1756 AAGCGGTGAGACTGGGAGACTGGAGCCAGATAAAGGAGGAGTGTGTCAGCAAGCAG 1815  
1267 CGTATTTGAAGGCGCTGTGCAAGATGACCGCGGAAGATGTATCAGGGAGTCTTTGC 1326  
1816 CAAGTTCAAGAGTGTGTGATCAACAGTATGAGAGAGAGTGTGACACAGCTGTGC 1875  
1327 CCAGTCTCCCTCTCTGAGAGAAATCTGTGAAATTCCTGCGAAATGGAGTGTGTG 1386  
1876 AGAGATGCCATTTCCCATCTGTCGCTGTGATGCCCATGCCCCATGCGGAAAGACTGTGTG 1935  
1387 CTGAGCGAGTGGAGGAGTGTCTATCTGTTCCAGTCTCTGTTCAAAATAAAATCAGAT 1446  
1936 CTCAGCATGTTCTACGTGTCTCTGCTCACACACTGCTCAGGGAAAAACGACAGAA 1995  
1447 GGGAAACAGACCGAGTCAAGAACTATCTGTGC---ACTGTGCTGGGAAAGTGGAAAGCCA 1503  
1996 GGGAAACAGATACGAGCAGCATCTATCTGTGCTATGCGGTGAAGAGTGGAAATGCG 2055  
1504 TGTCTCCCTCTAGCTGCTCTCAAGAGCATGTTTGTGTAATGACATTTCTGTATGCGAG 1563  
2056 TGTCCAAATAGAGTGTCTTCAAGAGAGTACGAAGCTGTAAAGAGTCTCTGTGACAGT 2115  
1564 CTTCACTGGGAGACATGCTGCTGGGCGCTTGTCTGAGGACACATTTGGTAACTGGCCCTT 1623  
2116 TACCACCTGGCAACTGTGCTCTGGGCGAGTGCATTTGAGGACACTCAGTATCTGTCTTC 2175  
1624 AATGCAACCATTTGGCTGGAATGGAGAGCCAGTGTGTGTAGGCATTTAGATCTGAGGA 1683  
2176 AACACAACTACGACTTGAATGGGAGCGCTCTGCTCTGTGCGCATGCGACAGAAAAA 2235  
1684 GTCCTCTGTCAAGAGTCACTGTGGGACAGTAATGACAAAGATGTCCAGATTTCTACT 1743

2236 GTCATCTGTGTGCGAGTCAATGTGGGCAAGTGGGACCCCAAAAATGTCTCTGAAGCCTT 2295  
1744 CGACTGAACTGTGCGCCCTGTTTCTCCATGCAAAAAGACTGTATTTGTGACTGCT 1803  
2296 CGACTGAACTGTGAGCCCTTGTCTGCTTCTTGTAAAGAGACTGTATTTGTGACCCCA 2355  
1804 TTCACTGAGTGGACACCCCTGCGCAAGGATGTGCAAGCAGGAAATGCCAGT---AAAA 1860  
2356 TATAGTGACTGACCATCATGCGCCCTCTTCTGTGTAAGAAAGGGGACTCCAGTATCAGGAAG 2415  
1861 CAGTCTCGATACAGAAATCATCATCCAGAGAGCCCAATGAGGCGCAGGAATGCCAGAT 1920  
2416 CAGTCTAGGCACTGGGTCTCATCTCAGCTGCCCAACCGGGGCCGAGACTGCGACAGAT 2475  
1921 ACCTTATATGAGGAGAGAGAGTGTGAAGATGTTTCTTGTGTCTGTATATTCGGTGAAG 1980  
2476 CCCCTCTATGAAGAGAGGCGCTGTGAGGCACCTCAAGCGTGCACAAAGCTACAGGTGGAAG 2535  
1981 CCACAGAAATGGAGCCCTTGCATCTTAGTGCAGAGTCTCTCTGGCAGGGAATAAACGGC 2040  
2536 ACTCACAATGGCGAGATGCCAATTAGTCCCTTGGAGCGTGCAACAGACAGCCCTTGA 2595  
2041 AGCAGTGAAGCCTGTGGAAGAGGGGTACAAACAAAGAGCTGTCTCATGCATCTCTGTATGAC 2100  
2596 GCACAGAGAGCTGTGGCCTGGCGGACAGCAAGAGCCATTACTTGTGCAAGCAAGAT 2655  
2101 AACCGTCAAGCAGAAATGATGAATGCTCAAGCAGACAAACGCGCATGCTCTCTTGTG 2160  
2656 GGAGACAGGCTGGAATCCATGAGTGCCTACAGTATGCAAGCCCTGTGCCAGCCCTTACC 2715  
2161 CAAGAATCCACAGTCCCACTGTGAGAGAGCTGACCTTCACTGCTGTGTCAGAGTTTACG 2220  
2716 CAGGCTGCGAGATCCCTGCGCAGATGACTGTCAATTGACAGCTGCTCAAGTTTCT 2775  
2221 CCCTGCTCCACAGATGTGAAGCCACAAAAAGTGTAGCGGGGACAGCTCAGGGAAGAGC 2280  
2776 TCATGCAATGGAGACTGTGTGTGCACTTAGGACACAGAAAGCGCACTCTTGTGGAAGAGT 2835  
2281 AGAAGAGAGAGAAATGCGCAGGATTTGACCTTTACCTCTAGTGGAGACAGAACTATGT 2340  
2836 AAAAAGAGGAAATGTAATAATTTCCATTTGTATCCCTGATGAGACTCAGTATGT 2895  
2341 CTTTGTGATGAATTTATATCCAACTTTATGGAACTGGTCAGATTCGATTTCTCCAGAA 2400  
2896 CTTTGTGACAAATATAATGCAACCTGTGGGAACTGGTCAGACTGTATTTTACCAGAG 2955  
2401 GCGAGAGGAGCCCTCACCGAGGACTGGGTAACAGCAGACAGCAAGAAATGTGGAGAA 2460  
2956 GGAAGAGTGGAAAGTGTGCTGGGAATGAAAGTACAGAGAGACATCAAGGAATGGGACAA 3015  
2461 GGCCTGCGCTTTTCGAGCAGTAGCTGTTCTGATAAAAATGGAAGACCTGTTGACCCCTCC 2520  
3016 GGATATCTTACCAAGCAATGSCATGCTACGATCAAAATGSCAGGCTTGTGGAACAATCT 3075  
2521 TTCTGCGAGAGCTGTGTTTACATTCAGAAATAATGTGTCTATCTCTGCCCCATTTGATGTC 2580  
3076 AGATGTAACAGCCATGTTTACATTTGAGGAGCGCTGCATCATCCCTGCGCCCTCAGACTGC 3135  
2581 AAGTTAAGCGATGTTGTCTGATTTGGGGTCTTGTGAGTTTCTGAGTTTCTGGAATTTGAGTGA 2640  
3136 AAGCTCAGTGTGTGCTCACTGTGCGGCTGCGAGAGTCTCTGTGGAGTGTGTGAAG 3195  
2641 ATTGATCCAAATGGCTAAAAAGAAACCTTACAAATGAGGAGCAAGCCATGTGCCAAATG 2700  
3196 GTTCGTTCTTAATGGCTGCTGMAAAACCATATAATGAGGAAGCGCTTGGCCCAAACTG 3255  
2701 GATC-----TCAAGAAATCAGTACATGAGCAGTCCCATGTTACAGTGTGCAATCAG 2754  
3256 GACCATGTCAACCGAGCAAGGTGTATGAGTGTGTCCTATGCCAGTGTGCAACCG 3315  
2755 TATTCTGCGTTGTAGAACACTGTTCTTTTCATGCAAAATCAACAATGAGCTGAGGTCCCTG 2814  
3316 TACCTATGGTCAAGAGCCCTGAGGAGCATCTGCAAGGTGACCTTTGTGAATATGCGGAG 3375

2815 CGCTGTGGAGGAGAACACAATCTAGAAAATCAGATGTGT---GAATATCTGGATGT 2871  
2872 GAAGGTGGACAGTGTAGTACCACTGTGCAACCAAGATGAATTCCTCCCAAGAACCCAG 2931  
2932 TCCTGTCTCTTATGTCTCCCAATGAGTGTGTCATCTGTGAGTGGGAGCTTTGGAGCAAA 2991  
2992 TGCCCAAGTGTAGTGTGCAATCTGAGGACTGTGTGATATCTGAATGGGTCCATGGACCCAA 3055  
3056 TGCTGTCTTGTGCAATCAAGACAGTTTCCGCAAAAGGTGAGTGTGATGCCATCAGACAA 3119  
3119 CCATCACTGAATCAAGGAGCTGTGTGTAAGACTCACAGTGTGAGTGTGCTCTCTGAT 3183  
3183 CCAGCTGTAGGAGAGATCTTCCCTATATGCTGTGAAAGAACCTGTAACTTGAAC 3247  
3247 GAAAAATGCTTCCAGTTCAGTACAATCTAACAGAGTGTGAGCAGATGCCAGCTGAGTGA 3311  
3311 AAAAACTGTCTACCACTATGATTAATGTAACTGAGTGTGATGTGAGTGTGAGTGTGAG 3375  
3375 AACGACCTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 3439  
3439 AAGGAGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 3503  
3503 AAGCCAGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 3567  
3567 AAGTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 3631  
3631 AAGTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 3695  
3695 AAGTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 3759  
3759 AAGTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 3823  
3823 AAGTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 3887  
3887 AAGTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 3951  
3951 AAGTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 4015  
3951 AAGTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 4079  
3951 AAGTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 4143  
3951 AAGTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 4207  
3951 AAGTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 4271  
3951 AAGTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 4335  
3951 AAGTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 4399  
3951 AAGTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 4463  
3951 AAGTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 4527  
3951 AAGTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 4591  
3951 AAGTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 4655  
3951 AAGTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 4719  
3951 AAGTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 4783  
3951 AAGTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 4847  
3951 AAGTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 4911  
3951 AAGTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 4975  
3951 AAGTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 5039  
3951 AAGTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 5103  
3951 AAGTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 5167  
3951 AAGTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 5231  
3951 AAGTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 5295  
3951 AAGTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 5359  
3951 AAGTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 5423  
3951 AAGTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 5487  
3951 AAGTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 5551  
3951 AAGTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 5615  
3951 AAGTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 5679  
3951 AAGTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 5743  
3951 AAGTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 5807  
3951 AAGTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 5871  
3951 AAGTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 5935  
3951 AAGTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 6000

3874 TACATCGGAAGCAAGCTCTTTGGAAACAATAACGAACAACTGTATGTCGACGCTTCA 3933  
4456 TATAAATGGATGGCCAGTGTGGAAGGGCTCTTCCGACACAGTGTGTCGAAAGTCA 4515  
3934 GATGCGGTAAATGTCAAGAGAGCTGTCCCTCAGCCCGTCTCTGCTGCTGCAATTCGGCAG 3993  
4516 GATGCTATAAATGTAAACAGGGGGCTGTGCTGATGATGAGCCGCTGATGCGCAGAGTCT 4575  
3994 TGCATTCAGGCTCGACAAAACCTTTCTCTACTGTATACACAGGGTGGAGTCTGTGGTTGT 4053  
4576 TGTAAACCCACGCTGTAGTCAACCCCACTCTGTACTGTAGCGAGACAAAACATGCCATGT 4635  
4054 GAGAAGGCTATACAGAGATAAAGAAATCAAAATGAAATGAAATGAAATGAAATGAAATG 4105  
4636 GAGAAGGCTATACAGAGATAAAGAAATGAAATGAAATGAAATGAAATGAAATGAAATG 4695  
4106 -----TGAAGTACAGGCTCAGAGGATAAAGAAATGAAATGAAATGAAATGAAATG 4157  
4696 CCGTGTGTGTATTTACCCACCATGGAGGACAAAAGAGAGATGTGAAAACCAAGTCGGGCT 4755  
4158 GAAAAACAGACCTGTGAAATTCAAAAATACATGATATTTTAAAGGA--TGGTCTCTTCAA 4215  
4756 GTACATCAACCCCAACCTCCAGTAACCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4815  
4216 CCAGTGTGATCCAGATGCGCGAGTAAAGAAATTTGGTGTATGCGGTTTCAGGTGGGCTTTT 4275  
4816 CCATTTGGGCGAGATGGAGACCTGAGACCTGAGGCTTTTACCGTGTAGCAGCTGGGCAATTT 4875  
4276 CTATCATGATTTTCTTAATTTACTTCTACTCTGTTTGCAGAGGAGGAGGAGGAGGAGGAG 4332  
4876 GTGTACTCATCTTTATTTGCTCCATGATTTATCTAGTTTGCAGAGGAGGAGGAGGAGGAG 4935  
4333 CATCAAGACACCTCCCAACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4392  
4936 CAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4995  
4393 ATGTAA 4398  
4996 ATGTAA 5001

## RESULT 6

US-10-037-270-332  
; Sequence 332, Application US/10037270  
; Publication No US20030104529A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Ren, Feiyan  
; APPLICANT: Chen, Rui-hong  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Yang, Yonghong  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Zhou, ping  
; APPLICANT: Ma, Yunding  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Wang, Zhiwei  
; APPLICANT: Tillinghast, John  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and  
; TITLE OF INVENTION: Polypeptides  
; FILE REFERENCE: 784CIP2B  
; CURRENT APPLICATION NUMBER: US/10/037,270  
; CURRENT FILING DATE: 2002-01-04  
; PRIOR APPLICATION NUMBER: 09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 1104

SOFTWARE: pt\_FL\_genes Version 1.0  
SEQ ID NO 332  
LENGTH: 6378  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (271)..(5037)  
NAME/KEY: misc feature  
LOCATION: (1)..(6378)  
OTHER INFORMATION: n = a,t,c or g  
JS-10-037-270-332

Query Match 28.0%; Score 1229.6; DB 14; Length 6378;  
Best Local Similarity 57.6%; Pred. No. 0;  
Matches 2596; Conservative 0; Mismatches 1784; Indels 126; Gaps 17;  
2Y 10 TGCATTGAGAGCTGACCGAATGTGGTTGCAATGAAATATATGCAACACTTGGCCCTT 69  
Db 541 TGCATCCAGAAAGACAAAGACATTCCTGGGAGGATATCATCTGTGAGTACTTTGAGCCC 600  
2Y 70 CAGCCTCCTACAGAACAGGCTTGCCTCATTCCTGTGCCCCGGGATTTGTGTATCTGAG 129  
Db 601 AAGCCTCTCTGGAGCAGGCTTGCCTCATTCCTGTGCCAGCAAGATTGCATCGTGTGAA 660  
2Y 130 TTCTTACATGGTCCAACTGTAGCAAGGATGTGGAGAGAAATTGACGATGAAGTCCG 189  
Db 661 TTTTCTGCTGTGCTCGAATCTCTCAAGACCTGGGAGGCTCCAGCAGCCGGAGCGGT 720  
2Y 190 GCGGTCTATAGTCCCTCTCTTTTGGTGGTTCGAATCTGCAATCTGACTGAGTCAAGA 249  
Db 721 CATGTGTGGGCCCCCGAGTTCGGAGGCTCTGGCTGTCCAACTGTACGGAGTTCCAG 780  
2Y 250 GCTGTGATGCTCCCAATTCCTGTCTTTGGGGAAGAGAAATATACATTTAGCCCTTAAG 309  
Db 781 GTGTG-----CCAACTCCATGCGAGCCGAGGAGCTCAGGTACAGCTTGCAT 831  
2Y 310 GTTGGACCATGGAGTAAATCAGACTGCTCATCTTAAAGAAATTAATCCAGCGGAAGA 369  
Db 832 GTGGGGCCCTGGAGCAGCTGCTCAATGCCCCATCTCCGACAGTAAGAAGCAAGGAGA 891  
2Y 370 ACTGTCTTGGATTTTAA-----CTCTGATTCAATGAG 402  
Db 892 CGCGGGAAGATAAAGAACGGGAAAGACCGCAGCAAGAGGTAAGGATCCAGAGCC 951  
2Y 403 CGAGTCACTTTAAACATCAAGTTACA-----AAGCATCATCATTCGAAGTCT 453  
Db 952 CGCGAGCTTTAAGAAAGAGAGAAACAGAAACAGGACAGCAAGCAAGAGAACAAATAT 1011  
2Y 454 TGGCAATAGAGATAGGTATCAAAACCGGAGGTTTCGTGTACAAGAAAGTATGACAA 513  
Db 1012 TGGACATCAGATTGGATATCAGACAGAGAGGTTATGTGCATTAAAGACGGGGA 1071  
2Y 514 AATGCTATGTAAGCCTTTGCTTCAAGATTCTTCCCATGATGACTGTTGAGTCTGATC 573  
Db 1072 GCTGCTGATTAAAGCTTTTCCAGCAAGAGAAAGCTTCCATGACCTTCCAGTCTCTGTG 1131  
2Y 574 ATGCCCAAGACTGTGAACCTCCAGTGTCTCTTGGAGCCCTGCTCCAGACATGC 633  
Db 1132 ATCAACAAAGTGCAGGTTTCGAGTGTGAGAGTGGAGCCCTGCTCAAAAACATGC 1191  
2Y 634 CGTTTACGGAGTCTCTTTGCCAGGATTTAGAGCAGGAGCCGGAACGTGAAGCAATGGCT 693  
Db 1192 CATGACATGGTGTCCCTGCGGACATCTGTGTAAGGACACGAACCATCAGGAGTTTCCC 1251  
2Y 694 ATTGGAGTGGAGAGTGTCTGACTTCTTGAAGAGAGCCCTGATTTGTAAGA 753  
Db 1252 ATTGGCAGTGAAGAGGATGTCAGAAATTTGAAGAAAGAAACCCCTGTTGTCTCAAGGA 1311  
2Y 754 GAACCTTCTGAGCAATGCTCCAGGATATCTCTGGAGAACTTCTGAATGGAAGAAATGCCAA 813  
Db 1312 GATGGAGTTGTCCCTGTGCCAGTATGGCTGGAGAACTACAGAGTGGACTGAGTGGCT 1371

814 GTCTCTCTCTCTCTCGAGCAGCAGGATCCCACTGGCATGTGACGGACCCGCTGTGTGGC 873  
Db 1372 GTGGACCTTTTGTCTAGTCAGCAGGACAGAGGCGGCAACAGACGGCCCTCTGTGGA 1431  
2Y 874 GGTGGATCCAGACCCCGGAGGTGTACTGTGCCAG-----AGCGTACAGCA 921  
Db 1432 GGGGCGATCCAGACCCCGAGAGGTGTACTGTGCGTCAGGCCAACGAAACCTCTCTCACAA 1491  
2Y 922 GCTCGCCACTGTAGGGCCCAAGGAAGTCTCTAGACCTGTGTGAAAGGCAATTATGTGGGA 981  
Db 1492 TTAAGTACCACAAGAACAAAGAGCCTCAAGCGCAATGACCTTAAATATATGCACTGGA 1551  
2Y 982 CCGCCCGCTGTGCTCTAGCTCTGCAATATACCTTGTCTTACGAGTCAATAGTATCT 1041  
Db 1552 CCTATCCCTAATATACACAGCTGTGCCAATCTTGTCCAACTGAATGTGAAGTTTCA 1611  
2Y 1042 TCTGTGTGAGCTGGGCGCTGTGCATCCATGAAACCTGATGAACCTCAAGCGGGAAGAAA 1101  
Db 1612 CCTTGGTCAGCTTGGGACCTTGTACTTATGAACCTGTATATGATCAGCAAGGGAAGAAA 1671  
2Y 1102 GGATTTAGAACGAGCGACG-----CCATGTCTCTCATGGAATCTACAGGCGCTGCA 1152  
Db 1672 GGCTTCAAACTGAGGAAGCGCGCATTTACCAATGAGCCCACTGGAGGCTCTGGGGTAACC 1731  
2Y 1153 GGGCATTTGCCCTCATTTGGTGGAGTCTGTTCTGTGAGGATCCAATGTCTACCGATGG 1212  
Db 1732 GGAACCTGCCCTCACTTACTGGAAGCCATTCCTCTGTGAAGAGCCTGCTCTTATGACTGG 1791  
2Y 1213 CTGCA-----CAGAAAGGATCTGTTTCCCTGATCATGGA-----AATGTGCGCTGGGACAT 1266  
Db 1792 AAGCAGTGAAGCTGGGAGACTGCGAGCCAGATAAAGGAAGAGTGTGTCCAGGCAAG 1851  
2Y 1267 CGTATTTGAAGGCGCTGTGCAGAAATGACGGGGAGAGATGTATCAGGAGTCTTTGC 1326  
Db 1852 CAAGTTCAAGAGGTTGTGTGCATCAACAGTGTGGAAGAAAGTTGACACAGACTGTGC 1911  
2Y 1327 CCAGTTCCTCTCTGAGAGGAGTCTTGTGAAATTCCTGCCGAATGAGTGTGTG 1386  
Db 1912 AGAGATGCCATCTTCCCATCTCTGCTGTGCTGTGATGCCCTGCGGAAAGACTGTGTG 1971  
2Y 1387 CTGAGCGAGTGGACGGAGTGTCTATCTGTTCAGTCTCTGTTCAAAATAAAATCAAGAT 1446  
Db 1972 CTCAGCATGCTGTAGTGTCTCTCTGTCTCACACACCTGCTCAGGGGAAAGACAGAGAA 2031  
2Y 1447 GGGAAACAGACCAAGTCAAGAACTATCTGSC-----ACTGCTGGGAGGCTGGAAGCCCA 1503  
Db 2032 GGGAAACAGATACAGACAGATCCATCTCTGCGCTATCGGGGTGAAGAGTGGAAATTCG 2091  
2Y 1504 TGTCCCTCTAGTCAAGGCTCTCAAGAGCATCTGTTGTATGATCACTTCTCTGTATGAG 1563  
Db 2092 TGTCCAAATGACAGTCTTGTGCAAGAGTACGAAGCTGTATGAGCATCTCTGACAGTG 2151  
2Y 1564 CTTCACTGGGAGCATCGCTTGGGCGCTTGTCTGAGGACATTTGGTAATGCGCCCTT 1623  
Db 2152 TACCACCTGGCAACTGTGCTGCTGGGCGCAGTGCATTTGAGGACACTCAGTATGCTCTTC 2211  
2Y 1624 AATGCAACCAATGCTGGATGGAAGAGCCAGTGTGCTGTAGGCAATTCAGACTCGGAGA 1683  
Db 2212 AACACACTACAGCTTGGMATGGGAGGCGCTCTGCTCTGTCGCAATGACAGCAAGAAA 2271  
2Y 1684 GTTTCTGTGTCAAGAGTCAAGTGGGCAAGATATGACCAAGAGATGTCCAGATTTACT 1743  
Db 2272 GTCATCTGTGTGCGAGTCAATGTGGGCGCAAGTGGGACCCCAAAATGTCTCTGAAGCCTT 2331  
2Y 1744 CGACTGAAACTGTGCGCCCTGTTTCTCCATGCAAAAGAAAGCTGATTTGTGACTGCT 1803  
Db 2332 CGACTGAAACTGTGAGGCTTGTCTGCTTCTTGTGAAGAGACTGTATTTGTGACCCCA 2391  
2Y 1804 TTCAGTGTAGTGGACACCTGCCCCAAGGATGTGCAAGCAAGAAATGCAAGT---AAAA 1860  
Db 2392 TATAGTGTAGTGGACATATGCCCTCTTCTGTGTAAGAAAGGGGACTCCAGTATCAGGAG 2451  
2Y 1861 CAGTCTCGATACAGATCATCATCCAGAGACGACCAATGGAGGCCAGGAATGCCAGAT 1920





QY	4054	GAGAAAGGGCTATACAGAGATPAATGAATAATCAATGGTTTCTCGATTACTGCA-----4105
Db	4672	GAAGAAGGGTACACTGAAGTCATGTCTCTTAAACAGCACCCCTTGAGCAATGCACACTATC 4731
QY	4106	-----TGAAGCTACCAGGCTCAGAGGATAAAAAGCTCATGTGAAAAACCTTCTG-G 4157
Db	4732	CCCGTGGTGGTATTTACCCACATGGAGGCAAAAGAGGAGATGTGAAAAACACGTGGGGCT 4791
QY	4158	GAAAAACAGACCTGTGAAATTCAAAAATPACATGATATTTTAAAAGGA--TGGTCTCTTTCAA 4215
Db	4792	GTACATCCACACCAACCCCTCCAGTAACCCAGCAGGACGGGGAAGGACCTGGTTTCTACAG 4851
QY	4216	CCACTTGATCCAGATGGCCGAGTAAAAATTTGGGTTTATGGCGTTTCAGGTGGCGCTTTT 4275
Db	4852	CCATTTGGGCCAGATGGGAGACTAAAGACCTGGGTTTACGGTGTAGCAGCTGGGGGCACTT 4911
QY	4276	CTCATCATGATTTTCCCTAATATTACTTCTCACTGTTTTCAGAGAGCCA---AAACCA 4332
Db	4912	GTGTTACTCATCTTTATTTGTTCTCCATGATTTATCTAGCTTTCGCAAAAGGCCAAAGAACCC 4971
QY	4333	CATCAAGGACACCTCCCAACAGAGCCCTCTGACCTTAGCCCTACGATGAGACTTAGAC 4392
Db	4972	CAAGAAGGCAAAACACCCGACTGAACCTTTAACTTAGCCTATGATGAGATGCCGAC 5031
QY	4393	ATGTAA 4398
Db	5032	ATGTAA 5037

```

RESULT 7
US-10-117-722-332
; Sequence 332, Application US/10117722
; Publication No. US20030219744A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20030219744A1el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2BCIP
; CURRENT APPLICATION NUMBER: US/10/117,722
; CURRENT FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1104
; SOFTWARE: BL-FL_genes Version 1.0
; SEQ ID NO 332
; LENGTH: 6378
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (271)..(5037)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(6378)
; OTHER INFORMATION: n = a,t,c or g
US-10-117-722-332

```

QY	70	CAGCCTCCTACAGAA	CAGSGTTGCCTCA	TTCTTGTCCCGGATTTG	TGATGATCTGAG	129		
Db	601	AAGCCTCTCTGGAG	CAGAGCTTTGCCTCA	TTCTTGTCCCGGATTTG	TGATGATCTGAG	660		
QY	130	TTCTTACATAGTGT	CCAACTGTACAGAG	GAATGTGGGAGAAAT	TGCAGCATAGAACTCGC	189		
Db	661	TTTTCTGCCTGGT	CCGAATGCTCCAG	AGCTTCGCGACGGGCT	CCAGCACCGGACGCT	720		
QY	190	CGGTCATAGCTCG	CCCTCTCTTTGGT	GTGTTGCAATGTCCAA	TCTGACTGAGTCAAGA	249		
Db	721	CATGTGTGGCGCC	CGGAGTTCGAGGCT	CTGGCTGTCCAAACCT	GACGGAGTTCAG	780		
QY	250	GCCTGTGATGCT	CCCATTTCTCTGT	CTCTTGGGAGAGGAAT	TACATTTTAGCTTTAAG	309		
Db	781	GTGTG-----	CCAATCCAGTCC	ATCGCAGGCGGAGGCT	CAGGTACAGCTCGAT	831		
QY	310	GTTGGACATGGAG	TAAATGCGACGCT	CATCTTAAGAAATTA	TATCCAAAGCGGAGA	369		
Db	832	GTGGGCGCTGGAG	CACTGCTCAATGCCCC	ACTCCGCAAGTAAGA	CAAGCAAGAGAGA	891		
QY	370	ACTGTTCTGGAT	TTTAA-----	-----	CTCTGATTTCAAATGAG	402		
Db	892	CGCGGAGAATAA	GAAACGGGAAAGG	ACCGCAGAAAGGAG	TAAAGGATCCAGAGCC	951		
QY	403	CGAGTCACCTTT	AAACATCAAGTT	TACA-----	AAACACATCATTTGAGTCT	453		
Db	952	CGCAGGCTTATT	TAAGAAAAGAGAA	CAAGAAACAGGCA	GAACAGCAAGAGAA	CAATAT	1011	
QY	454	TGGCAATAGAGAT	TAGGTTATCAAC	CCCGCAGGTTTCTGT	TACAAGAAGTGATGGCA	513		
Db	1012	TGGACATCCAG	ATTGGATATCAG	ACAGAGAGGTATGT	TGCTTAAACAGACGGGAAA	1071		
QY	514	AATGCTATGTAT	GACCTTTCCTTCA	AGATTCCTTCCCATTTG	ACTGTTTCACTCTGCATC	573		
Db	1072	GCTGCTGATTT	TAAGCTTTTTC	GCACGAGAGAGCTT	CCAATGACCTTCAGTCTCTGTG	1131		
QY	574	ATGCCAAAGACT	GTGAACCTCCCA	CTGCTCTCGAGCCCT	GTGTCCAAGACATGC	633		
Db	1132	ATCACCAGAGT	GCCAGGTTTCCG	AGTGTGTCAGAGTGG	AGCCCTGTCTCAAAAACATGC	1191		
QY	634	CGTTCAGGAGT	CTCTTGCAGAG	ATTTAGAGCAGSAG	CCGAGCTGAAGCACATGGCT	693		
Db	1192	CATGACATGTT	GTCCCTGCGAC	ACTCTGTAAAGA	CACGAAACCATCAGGAGTTTCCC	1251		
QY	694	ATTGGAGTGAA	AGGAGTGTCTG	AACTTCTTGAGAA	AGGCGCTGATTTTGAAGGA	753		
Db	1252	ATTGCGAGTGA	AAAGAGTGTCC	AGAAATTTGAAGAA	AAAGAACCTGTCTCTCAAGGA	1311		
QY	754	GAACTTTCGAG	CAATGTCCCGG	TATTCCTCGAGAA	CTTCTGAAATGGAAGAAATGCAA	813		
Db	1312	GATGAGTTGTC	CCCTGTGCC	ATGTGCTGTGAG	AACTACAGAGTGAAGTGGCGT	1371		
QY	814	GTCTCTCTCT	CTCGACAGCAG	AGATCCCCCACT	GTGCATGTGACGGGACCGGTGTGGC	873		
Db	1372	GTGACCTTTG	CTCAGTCAGC	AGGAGTACTGTGCCC	AG-----AGCTACCA	921		
QY	874	GGTGGATCCAG	ACCGCGGAGGT	GTACTGTGCGG	CCCAAGCCAAATGGA	CTTAAAAATATGCACTGGA	1431	
Db	1432	GGGGCATCCAG	ACCCGAGAGGT	GTACTGTGTCAG	GGCCCAACGAAACCTCTCTCACAA	1491		
QY	922	GCTCGCCCAT	GTGGGCCAAG	GAAGTCTCT	PAGA	CTGTGTGAAAAAGGCAATATGTGTGGGA	981	
Db	1492	TTAAGTACCA	CACAGAAACA	AGAGGCTCA	AAAGCCAAATGGA	CTTAAAAATATGCACTGGA	1551	
QY	982	CCCGCCCGTT	CGCTCTCAG	CTCTGCA	TATCCCTGTCTTAC	GGACTGCACATGATCT	1041	
Db	1552	CTATTCCT	TAATATA	CACAGCTGT	GCACATCTTCTGT	CCAACTGGAATGTGAAGTTCA	1611	
QY	1042	TCCTGGT	CAGCTCGG	GCCTGTG	CAATCA	TGAAACCTGTCA	TGAACCTTCAGGGGAAAAA	1101
Db	1612	CCTTGGT	CAGCTTCGG	GACCTGT	ACTTATGA	AACTGTAAATGAT	CAGCAAGGGGAAAAA	1671

1102 GGATTTAGACGAGCAGCG-----CCATGTCCTCATGAATCTACAGGCGCTGCA 1152  
1672 GGCTTCAAAGTGAAGAGCGCGCATTAACAATGAGCCCACTGGAGGCTCTGGGTAAACC 1731  
1153 GGCATTTGCCCTCATTTGGTGGAGTCTGTTCCCTTGTGAGGATCCAAATGTTGCTACCGATGG 1212  
1732 GGAAACTGGCCCTCACTTACTGGAAGCAATCCCTGTGAAGAGCTGCGCTTTATGACTGG 1791  
1213 CTGGCA--TCGAAAGGATCTGTTTCCCTGATCATGGA--AAATGTCGCCCTGGAGCAT 1266  
1792 AAGCAGTAGAGACTGGAGACTCGAGACAGATACGGAAGAGAGTGGTCCAGGACAG 1851  
1267 CGTATTCTGAAGCCCGCTGCGCAGATGACCGGGAGAGATGATACAGGAGTCTTTGCG 1326  
1852 CRAAGTTCAAGAGGTTGTGTGCATCAACAGTGATGGAGAGAGTTGACAGACAGCTGTGC 1911  
1327 CAGTTTCCCTCCTCTCTGAGAGGAGTCTTGTGAAATTCCTGCGCAATGACTGTGTG 1386  
1912 AGAGATGCCATCTTCCCATCCCTGTGGCTGTGATGCCCCCATGCCCGAAGACTGTGTG 1971  
1387 CTGAGCGAGTGGACGAGTGGTCACTCTGTTCCTCAGTCTCTGTTCAAAATAAAACTCAGAT 1446  
1972 CTGAGCACATGTTCTAGTGTCTCTCTCTCACACACTGCTCAGGGAAGACGACAGAA 2031  
1447 GGGAAACAGACAGGTCAGAAATATCTCTGGC---ACTGGCTGGGGAAGGTGGAAAGCCA 1503  
2032 GGGAAACAGATACGAGCAGCATCCATTTCTGGCCTATGCGGGTGAAGAGGTGGAATTCGC 2091  
1504 TGTCCCTCTAGTCAGGCTCTCCAAAGAGCATGTTGTGTAATGACCATTCCTGTATGCGAG 1563  
2092 TGTCCAAATAGCAGTCTTTCGAAAGTACGAGCTGTGTAAGAGCTGTTGACAGTG 2151  
1564 CTTCACTGGGAGACATGGCTTGGGCGCTTGTCTGAGGACACATTTGTTAACTGCCCTT 1623  
2152 TACCACTGGCAAACTGGTCCCTGGGCGCAAGTCAATGAGGACACCTCAGTATCGTCTTC 2211  
1624 AATGCAACCATTTGGCTGGAATGAGAGGACCACTGTTGTTAGGCAATTCAGACTCGGAGA 1683  
2212 AACCAACTACGACTGGAAATGGGAGGCTCTCTGCTGTGCGCATGAGACAGAAAGAA 2271  
1684 GTCTTCTGTCTCAAGAGTCACTGGGACAAAGTAATGACCAAAAGATGTCAGATTTCTACT 1743  
2272 GTCATCTGTGTGGAGTCAATGTTGGCCCAAGTGGGACCCCAAAAAATGTCCTGAAGCCTT 2331  
1744 CGACCTGAAACTGTGGCGCTCTTTTCTCCCATGCAAAAAGACTGTAATGTGACTGCT 1803  
2332 CGACCTGAAACTGTGAAGGCTTCTGCTCTCTTGTGAAGAGACTGTAATGTGACCCCA 2391  
1804 TTCAGTGAAGTGGACACCTGCCCCAAGGATGTGCCAAGCAGGAAATGCCACAGT---AAAA 1860  
2392 TATAGTCACTGGACATCATGCCCTCTTCTGTGTAAAGAGGGGACTCCAGTATCAGGAG 2451  
1861 CAGCTCTGATACAGAAATCAATCCAAAGAGAGAGCCAAATGGAGGCGCAGGAATGCCAGAT 1920  
2452 CAGCTTAGGCATCGGCTCATCAATCAGCTGCCAGCCCAACGCGGCGCGAGACTGCACAGAT 2511  
1921 ACCTTATAGAGAGAGAGAGTGTGAAGATGTTTCTCTGTCTCTATATCGGTGGAG 1980  
2512 CCCTCTATGAAGAGAGGCTGTGAGGACCTCAAGCGTGTCCAAAGCTACAGGTGGAG 2571  
1981 CCACAGAAATGGAGCCCTTGTCATCTTAGTCCAGAGTCTGTCTGGCAGGGAATAACGGGC 2040  
2572 ACTCAAAATGGCGCAGATGCCAATTTAGTCCCTTTGGAGCGTGCAACAGACAGCCCTGGA 2631  
2041 AGCAGTGAAGCTGTGGAAGGGGTTACAAACAGAGCTGTCTCATGATCTCTGATGAC 2100  
2632 GCACAGAGGCTGTGGGCTGTGGCGACAGGCAAGAGCCATTACTTGTGCGAAGCAAGAT 2691  
2101 AACCGGTACAGAAATGATGGAATGCTCAAGCAGACAAACGGCATGCTCTCTCTTGTG 2160  
2692 GGAGGACAGGCTGGAATCCATGAGTGCCTACAGTATGACGGCCCTGTGCGACCCCTTACC 2751  
2161 CAAGAAATGCACAGTCCCATGTGAGAGAGCTGCACCTTCACTGCTGGTCCAAAGTTTACG 2220

Db 2752 CAGCGCTCCAGATCCCTCCAGGATGACTGTCAATTGACAGCTGCTCAAGTTTCT 2811  
QY 2221 CCCTGTCCACGAACTGTGAAGCCACAAAAGTAGGGCGGACAGCTCA CAGGAAAGC 2280  
Db 2812 TCATGCAATGAGACTGTGGTGCAGTTAGACCAGAAAGCGCACTCTTGTGGAAAGT 2871  
QY 2281 AGAAGAGGAGAAATCCAGGATCTGACCTTTACCTCTAGTGGAGACAGAACTATGT 2340  
Db 2872 AAAAAAATAAATAATAAATTAATTTCCATTTGTATCCCTGATTGAGACTCAGTATTGT 2931  
QY 2341 CCTTGTGATGAATTAATATCCCAACTTATGGAACCTGTTGAGTGCATGCTCTTCCAGAA 2400  
Db 2932 CCTTGTGACAAATAATAATGCAACCTGTGGGAACTGTGGTCACTGATATTTTACCAGAG 2991  
QY 2401 GGCAGAGGAGGAGCTCACCGAGGATCGCGGTACAAGCAGACAGCAAGAAATGTGGAGAA 2460  
Db 2992 GGAAGATGGAAGTGTGTGGGAATGAAGTACAAGGAGACATCAAGGAATGGGACAA 3051  
QY 2461 GGCCTGGCTTTCCAGCAGTAGCCTGTTCTGATAAAATGGAAGACCTGTGACCCCTCC 2520  
Db 3052 GGATATCGTTACCAAGCAATGGCATGCTACGATCAAAATGCGAGGCTTGTGGAACATCT 3111  
QY 2521 TTCTGACAGACTCTGTTACATTCAGAAAAATGTGTCAATTCCTGCTGCCCTTGTGATGC 2580  
Db 3112 AGATGTAACAGCCATGTTTACATTTGAGGAGCCCTGTCATCATCCCTGCTCCCTCAGACTGC 3171  
QY 2581 AAGTTAAGCAGATTGGTCTAGTTGGGGTCTTGGAGTTTCACTTGTGGAATTTGGAGTGAGA 2640  
Db 3172 AAGCTCAGTGAGTGGTCCAACTGCTCGCGCTGCAGCAAGTCTGTGGAGTGGTGTGAAG 3231  
QY 2641 ATTGATTCCTAAATGGCTTAAAGAAAAACCTTACAAATGGAGAGACCAATGTCCCAAACTG 2700  
Db 3232 GTTCGTTCTAAATGGCTGCGTGAAGAAACCAATATAATGGAGGAAGGCTTGGCCCAACTG 3291  
QY 2701 GATC-----TCAAGAAATCAGGTACATGAGGAGTCCCATGTTTACAGTGAAGTCAATCAG 2754  
Db 3292 GACCATGTCAACAGGACAGGTGTATGAGTGTGTCCTCATGCCACAGTGAAGTGCACAG 3351  
QY 2755 TATTCCTGGTGTGAGAACACTGGTCTTTCATGCAAAATCAACAAATGAGCTGAGTCCCTG 2814  
Db 3352 TACCTATGGGTCAACAGAGCCCTGGAGCATCTGCAAGGTGACCTTTGTGAATATGCGGAG 3411  
QY 2815 CGCTGTGGAGGAGAAACCAATCTAGGAAATCAGATGTGT---GAATATCTGGGATGGT 2871  
Db 3412 AACTGTGGAGAGGCGGTGCAACCCGAAAGTGAAGTGAATGCAATGAGATGAGATGAGTGGC 3471  
QY 2872 GAAGGTGGAGCAGTGGATAGCAACCTGTGCAACAGGATGAAATTTCCCGAGAAACCCAG 2931  
Db 3472 CCTTCTGAACATGTAGAGGATTACCTCTGTGACCCAGAGAGATGCCCTGGGCTCTAGA 3531  
QY 2932 TCCTGTTCTTATGTGTCCTCAATGAGTGTGTCAATGCTCTGAGTGGGACTTTGGAGCAA 2991  
Db 3532 GTGTGCAAAATACCATGCCCTGAGGACTGTGTGATCTGAAATGGGGTCCATGGACCCAA 3591  
QY 2992 TGCCCAACAGTCAATCGGATCCCCACAAAT---CGAGAGAAAGAACTGCCACCTGCTAAGA 3048  
Db 3592 TGTGTTTGGCTTGCATCAAGCAGTTTCCGCAAGGTGAGTCAATCCCATCAGACAA 3651  
QY 3049 CCATCACTGAACCTCAAGACTTGTGTGAAGATCA CAGGTGAGCCCTTGCCTCTGAAAT 3108  
Db 3652 CCAGCTGATGAAGAAAGATCTTGGCCTTAATGCTGTTGAGAAAGAAACCTGTAACCTGAAC 3711  
QY 3109 GAAATTTGCTTCCAGTTCCAGTACAATCTAACAGGTGGAGCAGCATGCCAGCTGAGTGA 3168  
Db 3712 AAAAACTGTACCACTATGATTATATGTAACAGACTGGAGTACATGTCACTGAGTGAG 3771  
QY 3169 AACGACCCCTGTGTTGAAGGCTCAGGACCCGCTGCTTAAGCTGTGTGTGAGTGTGAGTGGC 3228  
Db 3772 AAGCAGTTTGTGAAATGGAATAAAAAAAGGATGTTGGATGTTGTTTGAAGTGTGAGTGGC 3831  
QY 3229 AAGCCAGTCACTGAGCAATGTGAGCAGCATAAATTTGGAGAGCCCGCAGAGAAATGAGC 3288

Db 3832 AAGTCAGTGCCTGAAATATTTGTGAAGCGCTTGGCTTGAGAGAACTGGCAGATGAAC 3891  
Qy 3289 ATTCCTCTGCTTGGTGAATGCTGCTCAACTGTCAGCTCTCAGGTGGACGGCTTCGACA 3348  
Db 3892 ACGTCTTGCAATGGTGGAAATGCCCTGTGAATGTCAGCTTCTGATGGTCTCTTGGTCA 3951  
Qy 3349 GAGTGTTCACAGACTGCTGGCCATGGAGGTGGAATGATGAGCCGGAATGATTTATCATATG 3408  
Db 3952 GAATGTTCTCAACATGCTGGCTCACAGGAAATATGATCCGAAGCAACAGTGAACCCAG 4011  
Qy 3409 CCAACCCAAAGGAGAGAGCCGATGCCACAGAGCTTACCCAGGAGAAACCTGCCCA 3468  
Db 4012 CCTTTCAAGGTGATGAGAGACCAATGCCCTTCTCCTGATGAGACAGTCCAAACCTGCCCA 4071  
Qy 3469 GTGACCCCTCTCAGCTGGTCTTGGCACTGGTCTGCATGTAAATTTGGAGGGTGA 3528  
Db 4072 GTGAAGCCTTGTATCGGTGCAATATGGCCAGTGGTCTCATGCCAAGTGCAGGAGGCC 4131  
Qy 3529 GACTGTGGGAGAGAGTTCAGATCCGAGCCTTCTGCGATGCTCCACAGTGGTCAATA 3588  
Db 4132 CAGTGTGGAGAGAGGACCGAACAAGGAACATTTCTTGTGTAGTAAGTGAAGGTGAGCT 4191  
Qy 3589 TCTCATGCACTGCAAGCTGTCGAGATGCACTGTGTGAGAAATGCCCTT----- 3638  
Db 4192 GATGATTCACAAAGTGGTGGATGAGGAATTTCTGTCTGACATTTGAATCATATAGAT 4251  
Qy 3639 -----TCAGGACAGATCCTGAAGCAGCTGTGTTCTGTGCTTGGCTTGGCCAGAGAGCTGCCAT 3693  
Db 4252 GGTAATAAAATATGGTTCTGGAGGAATCTCTGCAGCCAGCCTTGGCCAGGTGACTGTAT 4311  
Qy 3694 TTAACAGAAATGGTCAGAGTGAGACACATGTAAATTAACCTGCAATGATGAAGAAGCTTT 3753  
Db 4312 TTGAAGGACTGGTCTTCTGGAGCCTGTGTGCTGACCTGTGTGATGTGAGGATCTA 4371  
Qy 3754 GAGACTGTGGCCGCCAGTCTPAGATCAAGGACTTTTATAATTCAGTCTTTTGAAGAACCA 3813  
Db 4372 GGCTTTGGTGAATACAGGTGAGATCCAGACCGGTGATTTACAGAACTAGAGAATCAG 4431  
Qy 3814 GACAGCTGCCCCCAACAGGTTCTAGAAACAGCCCTTGTACAGAGGCAATGTTATCAC 3873  
Db 4432 CATCTGTGCCAGAGAGATGTTAGAAACAAATCATGTTATGATGACAGTGTCTATGAA 4491  
Qy 3874 TACACATGGAAGCAAGTCTTTGGAACAAATAACGAACTGTATGGTCCAGCCTTCA 3933  
Db 4492 TATAAATGGAATGGCAGTGTGGAAGGGCTTTTCCCGAACAGTGTGGTGTCAAGGTCA 4551  
Qy 3934 GATGGCGTTAATGTACAGAGGCTGTCTCCCTCAGGCCGCTGCTGCTGCTGCTGCTGCTG 3993  
Db 4552 GATGGTATAAATGTAAACAGGGGGTGTCTGTGTGATGAGCCAGCCTGTATGCCAGAGTCT 4611  
Qy 3994 TGCAATTCAGCCTGCAGAAACCTTTCTCTCTACTGTATACAGAGGTGAGTCTGTGTTGT 4053  
Db 4612 TGTAACCCACGCTGTAGTCAACCCCACTCGTACTGTAGCGAGACAAACATGCCATTGT 4671  
Qy 4054 GAGAGGGCTATACAGAGATTAATGAAATCAATGGTTTCTGGATTAAGTCA----- 4105  
Db 4672 GAAGAGGGGTACCTGAAAGTCAATGTCTTTCTAACAGCACCCTTGAGCAATGCACATTATC 4731  
Qy 4106 -----TGAAAGTACCAAGGTTCAGAGATATAAAGCTGATGTGAAAAACCTTTCTG-G 4157  
Db 4732 CCCGTGGTGTATTAACCCACCATGAGAGACAAAGAGAGATGTGAAAACCATGTCGGCT 4791  
Qy 4158 GAAAAACAGACTGTGAATCAAAAAATACATGATATTTTAAAGCA--TGCTCTTTCAA 4215  
Db 4792 GTACATCAACCCCAACCTCCAGTAACCCAGCAGGACGGGAGGAGCACTGGTTTCTACAG 4851  
Qy 4216 CCACCTTGTATCCAGATGGCCGAGTAAATTTGGGTTTATGGCTTTTACAGTGGCGCTTTT 4275  
Db 4852 CCAATTTGGCCAGATGGGAGACTAAAGACCTTGGGTTTACGGTGTAGCAGCTGGGCAATTT 4911  
Qy 4276 CTCATCATGATTTTCTTAATATTTACTTCTCACTACCTGTGTTTGCAGAGAGCCCA--AAACCA 4332  
Db 4912 GTGTTACTCATCTTTATTTGTTCTCCATGATTTATCTAGCTTCAAAAAAGCCCAAGAAACCC 4971

Qy 4333 CATCAAGACACACCTCCCAACAGAGCCTCTGACCTTAGCTAGCTAGGAGACTTAGAC 4392  
Db 4972 CAAGAAGGCAAAACACACCGACTGAACCTTTAACCTTAGCTATGATGAGATGCCGAC 5031  
Qy 4393 ATGTAA 4398  
Db 5032 ATGTAA 5037

## RESULT 8

US-10-104-047-958  
; Sequence 958, Application US/10104047  
; Publication No. US20030236392A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: No. US20030236392A1e1 full length cDNA  
; FILE REFERENCE: H1-A0105  
; CURRENT APPLICATION NUMBER: US/10/104,047  
; PRIOR FILING DATE: 2002-03-25  
; PRIOR APPLICATION NUMBER:  
; NUMBER OF SEQ ID NOS: 4096  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 958  
; LENGTH: 3053  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-104-047-958

Query Match 13.2%; Score 580.6; DB 15; Length 3053;  
Best Local Similarity 56.4%; Pred. No. 6.8e-178;  
Matches 1309; Conservative 0; Mismatches 934; Indels 78; Gaps 9;  
Qy 10 TCATTCAGAGCTGAACCGAAGTGTGTTGCAATGAATATGCGAACACATTTCGCCCTT 69  
Db 730 TGATCCAGAAAGACAAAGACATTCCTCGGAGGATATCATCTGTGAGTACTTTGAGGCC 789  
Qy 70 CAGCCTCTACAGACAGCTTGCCTCATTCCTTGTCCCGGATTTGTAGTATCTCAG 129  
Db 790 AAGCCTCTCTGGAGCAGCTTGCCTCATTCCTTCCAGCAAGATTGCTGCTCTGAA 849  
Qy 130 TTCTTACCATGGTCCAACTGTAGCAAGGATGTGGGAAGAAATTCAGCATAGAACTGC 189  
Db 850 TTTTCTGCTGTGCTCGAATGCTCCAAAGACCTGCGCAGCGGGCTCCAGCACCGGACGT 909  
Qy 190 GCGGTATAGCTCCCTCTCTTTGGTGGTTGCAATGTCCTCAATCTGACTGACTCAAGA 249  
Db 910 CATGTGGTGGCGCCTTCCGAGTTCGAGGCTCTGGCTGTCCAAACCTGACGAGTTCCAG 969  
Qy 250 GCCTGTGATGCTCCCATTTCTGTCTCTTGGGGAAGAGGAATATACATTTTAGCCTTAAG 309  
Db 970 GTGTG-----CCAATCCAGTCCATGCGAGCGGAGGAGCTCAGTACAGCTGCAT 1020  
Qy 310 GTTGACCATGAGTAATGACAGTGCCTCATCTTAAAGAAATTAATCCAGCGGAGA 369  
Db 1021 GTGGGCGCCTGGAGCACCTGTCTCAATGCCCACTCCGACAAAGTAAGCAAGAGGAGA 1080  
Qy 370 ACTGTTCTGATTTTAA-----CTCTGATTTCAAAATGAG 402  
Db 1081 CCGGGAAGATTAAGACGGGGAAGGACCGCAGCAAGGAGTAAGGATCCAGAGCC 1140  
Qy 403 CAGATCACTTTAAACATCAAAAGTTACA-----AAGCACATCATCATTCGAGTCT 453  
Db 1141 CCGAGCTTTATTAAGAAAAAGAGAAACAGAAACAGGCAAGACAGAACAGAGACAAATAT 1200  
Qy 454 TGGGCAATAGAGATAGGTATCAACACCGGAGGTTTCGTGTACAAAGATGATGACGACA 513  
Db 1201 TGGGACATCCAGATTGGATATCAGACCAAGAGGTTATGTGCAATTAACAGACGGGAAA 1260  
Qy 514 AATGCTATGTTAAGCCTTTGGCTTCAAGATTCTTCCCAATTCCTTCAGTCTCTGATC 573  
Db 1261 GCTGCTGATTTAAGCTTTTGGCAGCAAGAGAGCTTCCAATGACCTTCCAGTCTCTGTG 1320

1Y	574	ATGCCCAAAGA	CTGTGAAAC	TCCCOAGT	GGTCTCT	CTGTGAG	CCCTGTCT	CCAAGACATGC	633
1B	1321	ATCACAAAGAG	TGCCAGGTT	TTCCGAGT	GGTCAGAG	TGGAGGCC	CTGTCTCA	AAAAACATGC	1380
1Y	634	CGTTCAGGGAG	TCTCTTGC	CCAGAT	TTTACGAG	CAGGAG	CCGGAAC	GTGAAGC	693
1B	1381	CATGACATG	TGTCCTCT	CGAGCA	TCTGTAT	AGGACAC	GAACAT	CAGGAGTTTCCC	1440
1Y	694	ATTGAGGTGGA	AAAGAGTGT	CCTGAAC	TCTTGA	AAAGAG	CGCCTGCA	TTGTTGAAGA	753
1B	1441	ATTGGCAGTGA	AAAGAGTGT	CCAGAT	TTGAAGAA	AAAGAAC	CCCTGTTT	GTCTCAAGGA	1500
2Y	754	GRACITCTGCA	CAATGTC	CCAGTATT	CCTGGAG	BACTTCTG	AAATGGA	AAAGATGCCAA	813
1B	1501	GATGAGTTGT	CCCCGTGT	GCCACG	TATGGT	GTGAGAAC	TACAGAG	TGGAGT	1560
2Y	814	GTCTCTCTCT	CTCTGAG	CAGCAG	ATCCCA	CTGGCAT	GTGAC	GGGAC	873
1B	1561	GTGGACCC	TTTGTCT	CAGTCAG	CAGGAG	CAAGAG	CGCGCA	CCAGAC	1620
2Y	874	GGTGGATCAG	ACCCGGAG	TGTACT	GTGCCAG	-----	-----	AGCGTAC	921
1B	1621	GGGGGCATCC	AGACCCGAG	GTGTACT	CGTGCAG	CCACG	AAACCT	CTCTCACAA	1680
2Y	922	GCTCCGCACT	GAGGCCA	AGGAAGT	CTCTAG	ACCCTGTG	AAAAAG	CAATTATG	981
1B	1681	TTAAGTACCA	CAAGAAC	AAAGAG	CTCANAG	CCAATG	GA	CTTAA	1740
2Y	982	CCGGCCCG	TGCCCC	TCTCAG	CTCTGCA	TATCC	CTTGTCT	TACG	1041
1B	1741	CCATCCCTAA	TACTACTAC	AGCTGT	CCCACT	TCTTGT	TCCAACT	GAATGT	1800
2Y	1042	TCTGTGTAC	GTGGG	CTGTG	CACTCC	ATGAAAA	CTGTCA	TGAA	1101
1B	1801	CCTTGTGTAC	GTGGGA	CTTGTA	CTTATG	AAAACTGT	AA	TGTCAG	1860
2Y	1102	GGATTTAGA	ACGAGG	CAGCG	-----	-----	CCATGT	CTCTCAT	1152
1B	1861	GGCTTCAAA	CTGAGGA	CGCGCA	TATAC	CAATG	AGCCCA	CTGGAG	1920
2Y	1153	GGGCATTTG	CCCTCA	TTTGGT	GAGTCTGT	TCTTGT	GAGAT	CCATGT	1212
1B	1921	GGAACTG	CCCTC	ACTTACT	GTGAAG	CCAATCC	CTGTGA	AGAG	1980
2Y	1213	CTGGCA	---TCAGA	GGGATCTGT	TTCCCT	TGATCAT	TGGAA	---AATGT	1266
1B	1981	AAAGCAGT	GAGCTGG	GAAC	TGCGAG	CCAGATA	CGGA	AGAG	2040
2Y	1267	CGTATTTCT	GAGCG	CGTGTG	CCAGAT	ATGAC	CGGAG	AGAT	1326
1B	2041	CAAGTTCA	AGAGTTT	GTGTG	TCATCA	CAGTGT	GTGAG	AAAGTT	2100
2Y	1327	CCAGTTCCCC	CTCTCT	GAGAG	GAAGTCT	TGTG	AAATTC	CCCTG	1386
1B	2101	AGAGATG	CCATCTT	CCCCAT	CCCTGT	GGCCTGT	ATG	TGCC	2160
2Y	1387	CTGAGCGAG	TGGA	CGGAGT	GGTTCAT	CTGT	TCCAGT	CTCTGTT	1446
1B	2161	CTCAGCAC	ATGTTCT	AC	TGGTCT	CTCTG	CTCAC	ACCTGT	2220
2Y	1447	GGGAAACAG	ACAGGT	CAAGAA	CTATC	CTTGG	CACTGG	CTGG	1503
1B	2221	GGGAAACAG	ATA	CGACAC	ATCA	TCTGG	CTTATG	CGG	2280
2Y	1504	TGTCCCC	CTTACT	CAGGCT	CTCC	AGAG	CACTG	TTTGT	1563
1B	2281	TGTCCAA	ATAG	CAGTGT	TTTGC	AAAGT	ACG	AGCTGT	2340
2Y	1564	CTTCA	CTGGAG	ACATCG	CTTGG	GGCCCTT	GTCTGT	GAG	1623
1B	2341	TAC	CACTGG	CAAACT	TGGTCC	CTGGG	CCAGT	GCAT	2400

1624	QY	ATATGCAACCAATTGGCTGGAAATGGAGAGCCCACTGTGTGTGTGTAGGCATTCAGACTCGAGA	1688
2401	DB	AACACAACTACGACTTGGAAATGGGGAGGCTCTCTGCTCTGTGGCATGAGACAAGAAA	2460
1684	QY	GTCTTCTGTGTCAAGAGTCACTGGGACCAAGTAAATGACCAAAAGATGTCCAGATTCTACT	1743
2461	DB	GTCAATCTGTGCGAGTCAATGTGGGCCAAGTGGGACCCAAAATAATGTCTGAAGACCTT	2520
1744	QY	CGACCTGAACCTGTGTGGCCGCTGTTTTCTTCCCATGCAAAAAGACTGTATTGTGACTGCT	1803
2521	DB	CGACCTGAACCTGTAAAGGCTGTGTCTGCTTCTCTTTGTAAGAAGACTGTATTGTGACCCCA	2580
1804	QY	TTCAAGTAGTGGACACCCCTGCCCAAGGATGTGCCAAGCAGGAAATGCCACAGT---AAAA	1860
2581	DB	TATAGTAGCTGGACATCATGCCCCCTCTCGTGTAAAGAGGGGACTCCAGTATCAGGAAG	2640
1861	QY	CAGTCTCGATACAGAAATCATATCCAAAGACGACGCCAATGGAGGCCAGGAATGCCAGAT	1920
2641	DB	CAGTCTTAGGCATTCGGGTCTATCTTCAAGCTGCCAGGCCAACGGGGGCCGAGACTSCACAGAT	2700
1921	QY	ACCTTATATGAGGAGAGAGTGTGAAGATGTTTCTCTGTCTCTCTATATCGTGGGAAG	1980
2701	DB	CCCTCTATGAGAGAAAGGCTGTGAGGCACCTCAAGCGTGCCTAAGACTACAGTGGGAAG	2760
1981	QY	CCACAGAAATGGAGCCCTTGGATCTTAGTGCAGAGTCTGTCTGGCAGGGAATACGGGC	2040
2761	DB	ACTCACAAATGGCGCAGATGCCAAATTAGTCCCTTGGAGCGTGCACCAAGACACAGCCCTGGA	2820
2041	QY	AGCAGTGAAGCCTGTGGAAGGGGTTACAAAACAAGAGCTGTCTCATGCATCTCTGATGAC	2100
2821	DB	GCACAGAGAGGCTGTGGGCTGGCGCAGGCAAGAGCCATTACTTGTGCGAAGCAAGAT	2880
2101	QY	AACCGTCAAGCAAAATGATGGAATGCCCTCAAGCAGACAGAAAACGGCATGCTCTCTTGTG	2160
2881	DB	GGAGGACAGGCTGGAATCCATGAGTGCCCTACAGTATGACAGGCGCTGTGCCAGCCCTTACC	2940
2161	QY	CAAGATGACACAGTCCCATGTGAGAGAGAGCTGCACCTTCACTCTTGGTCCAAAGTTTACG	2220
2941	DB	CAGGCTCCAGATCCCTGCGCAGGATGACTGTCAATTGACAGCTGGTCCAAAGTTTCT	3000
2221	QY	CCCTGTCCACGAACTGTGAAGCCACAAAAGTAGCGCGG	2261
3001	DB	TCATGCAATGGAGACTGTGGTGCAGTTAGGACCAGAAAGCG	3041

RESULT 9

US-09-960-253-20

Sequence 20, Application US/09960253

Patent No. US20020123619A1

GENERAL INFORMATION:

APPLICANT: Benson, Darin R.

APPLICANT: Mohamath, Raodoh

APPLICANT: Lodes, Michael J.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER

FILE REFERENCE: 210121.556

CURRENT APPLICATION NUMBER: US/09/960,253

CURRENT FILING DATE: 2001-09-20

NUMBER OF SEQ ID NOS: 187

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 20

LENGTH: 516

TYPE: DNA

ORGANISM: Homo sapiens

US-09-960-253-20

	Query Match.	11.7%	Score 512.4;	DB 9;	Length 516;
	Best Local Similarity 99.8%;		Pred. No. 3.9e-156;		
	Matches 513;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
Qy	1713	AGTAATGACCAAAAGATGTCAGATTCTACTCGACCTGAACTGTGGCCCCCTGTTTCT	1772		
Db	3	AGTAATGACCAAAAGATGTCAGATTCTACTCGACCTGAACTGTGGCCCCCTGTTTCT	62		

QY 1773 CCATGCAAAAGACTGTATTGTGACTGCTTTTCAGTGAAGTGCACACCTTGCACAGGAT 1832  
Db 63 CCCATGCAAAAGACTGTATTGTGACTGCTTTTCAGTGAAGTGCACACCTTGCACAGGAT 122  
QY 1833 GTGCCAAGCAGGAAATCCACAGTAAACAGTCTCGATACAGAAATCATCATCCAAAGAGC 1892  
Db 123 GTGCCAAGCAGGAAATCCACAGTAAACAGTCTCGATACAGAAATCATCATCCAAAGAGC 182  
QY 1893 AGCAATGGAGGCCAGAAATCCAGATACCTTATATGAGAGAGAGAGTGTGAAGATGT 1952  
Db 183 AGCAATGGAGGCCAGAAATCCAGATACCTTATATGAGAGAGAGAGTGTGAAGATGT 242  
QY 1953 TTCCTTGTGCTCTATATCGTGGAAAGCCACAGAAATGAGCCCTTGCATCTTAGTGCC 2012  
Db 243 TTCCTTGTGCTCTATATCGTGGAAAGCCACAGAAATGAGCCCTTGCATCTTAGTGCC 302  
QY 2013 AGATCTGTCTGGCAGGAAATACCGGCAGCAGTGAAGCCTGTGGAAAGGGTTACAAC 2072  
Db 303 AGATCTGTCTGGCAGGAAATACCGGCAGCAGTGAAGCCTGTGGAAAGGGTTACAAC 362  
QY 2073 AAGAGCTGTCTCATGATCTCTGATGACAAACCGGTGAGCAAAATGATGAATGCTCAA 2132  
Db 363 AAGAGCTGTCTCATGATCTCTGATGACAAACCGGTGAGCAAAATGATGAATGCTCAA 422  
QY 2133 GCAGACAAACGGCAGTCTCTCTTGTGCAAGAAATGCACAGTCCCATGTGCAAGAGCTG 2192  
Db 423 GCAGACAAACGGCAGTCTCTCTTGTGCAAGAAATGCACAGTCCCATGTGCAAGAGCTG 482  
QY 2193 CACCTTCACTGCTTGTGCTCAAGTTTACGCCCTGC 2226  
Db 483 CACCTTCACTGCTTGTGCTCAAGTTTACGCCCTGC 516

RESULT 10

US-09-864-761-27172  
; Sequence 27172, Application US/09864761  
; Patent No. US2002048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aecmca-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662

; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 27172  
; LENGTH: 744  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC011231.1  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.89  
; OTHER INFORMATION: EST HUMAN HIT: BE295719.1, EVALUE 6.00e-53  
; OTHER INFORMATION: NT HIT: AF133643.1, EVALUE 1.90e+00  
; OTHER INFORMATION: SWISSPROT HIT: P35446, EVALUE 6.00e-10  
US-09-864-761-27172  
Query Match 10.5%; Score 460; DB 9; Length 744;  
Best Local Similarity 100.0%; Pred. No. 7.4e-139;  
Matches 460; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGTGCGCTGCATTCAGAGCTGAACCGAAGTCTGCTGCAAAATGAAATATGCAACAC 60  
Db 285 ATGTGCGCTGCATTCAGAGCTGAACCGAAGTCTGCTGCAAAATGAAATATGCAACAC 344  
QY 61 TTTTGCCCTTCAGCCTCTCTACAGAACAGCTTGCCTCATCTTCTCCCGGGATTGTGA 120  
Db 345 TTTTGCCCTTCAGCCTCTCTACAGAACAGCTTGCCTCATCTTCTCCCGGGATTGTGA 404  
QY 121 GTATCTGAGTCTTACCATGTGCTCAACGTAGCAAGGATGTGGAGAAATTCAGCAT 180  
Db 405 GTATCTGAGTCTTACCATGTGCTCAACGTAGCAAGGATGTGGAGAAATTCAGCAT 464  
QY 181 AGAATCTCGCGGGTCTATAGCTCCCTCTCTTTGGTGTGTTGCAATGTCCAAATCTGACT 240  
Db 465 AGAATCTCGCGGGTCTATAGCTCCCTCTCTTTGGTGTGTTGCAATGTCCAAATCTGACT 524  
QY 241 GAGTCAAGAGCTGTGATGTCTCCCATTTCTGCTCTTTGGGGAAGAGAAATATACATTT 300  
Db 525 GAGTCAAGAGCTGTGATGTCTCCCATTTCTGCTCTTTGGGGAAGAGAAATATACATTT 584  
QY 301 AGCCTTAAGGTTGGACCATGGAGTAATGCAGACTGCTCATCTTAAAGAAATTAATCCA 360  
Db 585 AGCCTTAAGGTTGGACCATGGAGTAATGCAGACTGCTCATCTTAAAGAAATTAATCCA 644  
QY 361 AGCGGAAGAACTGTTCTGGATTTTAACATCTCAATCAATGAGCGAGTCACTTTAAACAT 420  
Db 645 AGCGGAAGAACTGTTCTGGATTTTAACATCTCAATCAATGAGCGAGTCACTTTAAACAT 704  
QY 421 CAAAGTTACAAAGCAATCATCTTGAAGTCTTGGGCAA 460  
Db 705 CAAAGTTACAAAGCAATCATCTTGAAGTCTTGGGCAA 744

RESULT 11

US-10-029-386-5068/c  
; Sequence 5068, Application US/10029386  
; Publication No. US20030194704A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO

RESULT 14



US-10-029-386-22453/c  
 ; Sequence 22453, Application US/10029386  
 ; Publication No. US20030194704A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Penn, Sharon G.  
 ; APPLICANT: Rank, David R.  
 ; APPLICANT: Hanzel, David K.  
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G  
 ; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO  
 ; FILE REFERENCE: AEMICA-X-2  
 ; CURRENT APPLICATION NUMBER: US/10/029,386  
 ; CURRENT FILING DATE: 2001-12-20  
 ; NUMBER OF SEQ ID NOS: 34288  
 ; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
 ; SEQ ID NO 22453  
 ; LENGTH: 174  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; OTHER INFORMATION: MAP TO AC012000.3  
 ; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1  
 ; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.53  
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.93  
 ; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.81  
 ; OTHER INFORMATION: NT HIT: g14721097, EVALUOE 2.00e-93  
 ; OTHER INFORMATION: SWISSPROT HIT: P36297, EVALUOE 1.10e+00  
 ; OTHER INFORMATION: EST\_HUMAN HIT: BE940082.1, EVALUOE 8.90e-01  
 ; US-10-029-386-22453

Query Match 4.0%; Score 174; DB 14; Length 174;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-45;  
 Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1103 GATTAGAACGAGCAGCGCCATGCTCTCATGGAATCTACAGGCGCTGCAGGCGCATGGCC 1162  
 DB 174 GATTAGAACGAGCAGCGCCATGCTCTCATGGAATCTACAGGCGCTGCAGGCGCATGGCC 115  
 QY 1163 CTCATTGGTGAGCTCTGCTTCTGTGAGGATCCAAATGCTACCGATGCTGCGCATCAG 1222  
 DB 114 CTCATTGGTGAGCTCTGCTTCTGTGAGGATCCAAATGCTACCGATGCTGCGCATCAG 55  
 QY 1223 AAGGATCTGTTCCCTGATCATGGAATAATGTCGCTGGGACATCGTATTCTGA 1276  
 DB 54 AAGGATCTGTTCCCTGATCATGGAATAATGTCGCTGGGACATCGTATTCTGA 1

RESULT 15  
 US-09-919-172-96  
 ; Sequence 96, Application US/09919172  
 ; Patent No. US20020119463A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Paris, Mary  
 ; APPLICANT: Turner, Christopher M.  
 ; TITLE OF INVENTION: PROSTATE CANCER MARKERS  
 ; FILE REFERENCE: PA-0036 US  
 ; CURRENT APPLICATION NUMBER: US/09/919,172  
 ; CURRENT FILING DATE: 2001-07-30  
 ; PRIOR APPLICATION NUMBER: 60/222,469  
 ; PRIOR FILING DATE: 2000-07-28  
 ; NUMBER OF SEQ ID NOS: 102  
 ; SOFTWARE: PERL Program  
 ; SEQ ID NO 96  
 ; LENGTH: 579  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; OTHER INFORMATION: Incyte ID No. US20020119463A1 041764.1  
 ; US-09-919-172-96

Query Match 3.7%; Score 161.8; DB 9; Length 579;  
 Best Local Similarity 58.7%; Pred. No. 2.7e-41;  
 Matches 340; Conservative 0; Mismatches 227; Indels 12; Gaps 3;

QY 2662 GAAAAACCTTACATGAGGAGCAGCACTGTCCCAACTGGATCTCAAGATCAGGTAC-- 2719  
 DB 1 GAAAAACCATATAATGAGGAGGAGGCCCTTGGCCCCAACTGGACCAATGTCAACAGGCACAG 60  
 QY 2720 ----ATGAGGCAGTCCCATGTTTACAGTGAGTGCAATCAGTATTCTCTGGGTTGTAGAACAC 2775  
 DB 61 GTGTATGAGGTTGTCCCATGTCACAGTGACTGCAACCCAGTACCTATGGTGCACAGAGCCC 120  
 QY 2776 TGGTCTTCATGCAAAAATCAACAATCAGTGAGTGAGTCCCTGGCTGTGGAGGAGGACACAA 2835  
 DB 121 TGGAGCATCTGCAAGGTGACCTTTGTGAATATGCGGAGAACTGTGGAGAGGGCGTGCAA 180  
 QY 2836 TCTAGGAAAATCAGATGTGT---GAATACTCGGATGCTGAAGGTGGAGCAGTGGATAGC 2892  
 DB 181 ACCCGAAAAGTGAGATGATGCAGATACAGCAGATGGCCCTTCTGAACATGTAGAGGAT 240  
 QY 2893 AACCTGTGCAACAGGATGAAATTTCCCCCAGAAAACCCAGTCCCTGTTCTTATGTGTCCC 2952  
 DB 241 TACCTCTGTGACCCAGAGAGATGCCCTGTGGCTCTAGAGTGTGCAAAATTACCATGCCCT 300  
 QY 2953 AATGAGTGTGTCTGTCTGAGTGCGGACTTTGGAGCAAAATGCCACAGTCAATGCGATCCC 3012  
 DB 301 GAGGACTGTGTGATATCTGAATGGGTCCATGGACCCAAATGTGTTTGCCTTGCATCAA 360  
 QY 3013 CACACAAT---GCAGAGAGAACTCGCCACCTGTCTAAGACCATCACTGAACCTCAAGGACT 3069  
 DB 361 AGCAGTTTCCCGCAAGGTCAGCTGATCCCATCAGACAAACCCAGCTGATGAAGGAAGATCT 420  
 QY 3070 TGTGCTGAAGACTCACAGGTGCAGCTTGCCTTGCCTGAATGAAAATTTCTTCCAGTCCAG 3129  
 DB 421 TCCCTAATGCTGTTGAGAAAAGAACCTGTAACTGACAAAAACTGCTACCATATGAT 480  
 QY 3130 TACAATCTAACAGAGTGGAGCACATGCGCAGCTGAGTGAATAACGCCACCTGTGGTCAAGGC 3189  
 DB 481 TATAATGTAACAGACTGGAGTACATGTGAGTGAAGAGGCGAGTTTGTGGAATGGA 540  
 QY 3190 GTCAGGACCCCGCTGCTAAGCTGTGTGTCAGTGCATGCG 3228  
 DB 541 ATAAAAACAAGATGTTGGATTGTTCGAAGTATGCG 579

Search completed: February 24, 2004, 00:59:43  
 Job time : 1490.49 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

M protein - protein search, using sw model

Run on: February 23, 2004, 20:29:46 ; Search time 134.011 Seconds  
(without alignments)  
3088.802 Million cell updates/sec

Title: US-10-022-710-2

Effect score: 8241

Sequence: 1 MVRCTQKLNRTIVANEICEH.....OSTPPQKRLTLAYDGDLDL 1465

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_29Jan04.\*  
1: Geneseqp1980s.\*  
2: Geneseqp1990s.\*  
3: Geneseqp2000s.\*  
4: Geneseqp2001s.\*  
5: Geneseqp2002s.\*  
6: Geneseqp2003as.\*  
7: Geneseqp2003bs.\*  
8: Geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	8227	99.8	1574	7	ADE08053 Novel pro
2	4296	52.1	1551	4	AAM41081 Human pol
3	4296	52.1	1588	4	AAB20155 Secreted
4	4296	52.1	1588	6	ABR58456 Human NOV
5	4296	52.1	1588	6	ABR58456 Human NOV
6	4296	52.1	1624	5	AAB23979 Human LP2
7	4292	52.1	1588	4	AAM39295 Human NOV
8	4290	52.1	1588	6	ABR58450 Human NOV
9	4284	52.0	1588	6	ABR58459 Human NOV
10	4113.5	49.9	1545	6	ABR58453 Human NOV
11	4095.5	49.7	1549	6	ABR58455 Human NOV
12	4033	48.9	1490	6	ABR58454 Human NOV
13	2729	33.1	484	5	AAO15279 Human lar
14	2126	25.8	933	7	ADB64774 Human pro
15	1762.5	21.4	577	6	ABR58460 Human NOV
16	1756.5	21.3	571	6	ABR58452 Human NOV
17	1389.5	16.9	535	6	ABR58458 Human NOV
18	1389.5	16.9	535	6	ABR58457 Human NOV
19	1361.5	16.5	467	6	ABR58451 Human NOV
20	1332.5	16.2	617	3	AAB42496 Human ORF
21	1121.5	13.6	284	4	ABG05063 Novel hum
22	852	10.3	247	4	AAM70792 Human bon
23	852	10.3	247	4	AAM58317 Human bra
24	812	9.9	149	4	ABG05062 Novel hum
25	694	8.4	2150	3	AAV53898 Amino aci

26	694	8.4	2165	4	AAB90617 Human sec
27	674	8.2	1916	5	AAE19173 Human pro
28	673	8.2	1907	5	AAU77133 Human pro
29	673	8.2	1935	5	AAU72896 Human met
30	657	8.0	1882	4	AAU72286 Human ADA
31	655.5	8.0	4123	7	ABU62079 Human jel
32	646	7.8	1934	4	ABG72301 Human ADA
33	598.5	7.3	4561	4	ABG30203 Novel hum
34	598.5	7.3	9222	4	ABG21064 Novel hum
35	597.5	7.3	1629	5	ABG30703 Human agg
36	597.5	7.3	1629	5	AAO14448 Human ADA
37	578.5	7.0	1602	5	ABG30702 Human agg
38	547.5	6.6	307	6	ABR58461 Human NOV
39	507.5	6.2	874	3	AAO1431 Human TAN
40	498	6.0	1762	7	AAO30834 Human cel
41	495	6.0	1762	5	AAU9587 Novel hum
42	482.5	5.9	966	5	AAU80189 Human TSP
43	477.5	5.8	1745	4	AAU90551 Human sec
44	477.5	5.8	1745	5	ABG65473 Human alb
45	477.5	5.8	1766	4	ABB10246 Human CDN

## ALIGNMENTS

### RESULT 1

ADE08053  
ID ADE08053 standard; protein; 1574 AA.

AC ADE08053;

XX 29-JAN-2004 (first entry)

DT Novel protein (useful for identifying genetic disorders) #208.

DE novel gene; novel protein; tissue marker; molecular weight marker;

KW chromosome marker; genetic disorder.

XX Unidentified.

XX WO2003054152-A2.

XX 03-JUL-2003.

PD 10-DEC-2002; 2002WO-US039555.

XX 10-DEC-2001; 2001US-0339739P.

XX 11-DEC-2001; 2001US-0339453P.

PR 14-MAR-2002; 2002US-0365091P.

PR 14-MAR-2002; 2002US-0365384P.

PR 12-APR-2002; 2002US-0372381P.

PR 12-APR-2002; 2002US-0372815P.

PR 22-APR-2002; 2002US-00128558.

XX 24-APR-2002; 2002US-0376045P.

(HYSE-) HYSEQ INC.

XX Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao OA, Wang J;

PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang Z;

PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ;

XX WPI; 2003-569235/53.

DR N-PSDB; ADE07142.

XX New polynucleotides, useful for expressing recombinant proteins for

PT analysis, characterization or therapeutic use, or as markers for tissues

PT in which the corresponding protein is preferentially expressed.

XX Claim 20; SEQ ID NO 1119; 1177bp; English.

CC The invention comprises the amino acid and coding sequences of novel

CC proteins. The DNA and protein sequences of the invention are useful as:

CC markers for tissues in which the corresponding protein is preferentially

CC expressed; as molecular weight markers on gels; as chromosome markers or  
CC tags; to identify chromosomes or to map related gene positions; and to  
CC compare with endogenous DNA sequences in patients to identify potential  
CC genetic disorders. The present amino acid sequence represents a protein  
CC of the invention.  
XX  
XX

Sequence 1574 AA;

Query Match	99.8%;	Score 8227;	DB 7;	Length 1574;
Best Local Similarity	99.8%;	Pred. No. 0;		
Matches 1464;	Conservative 1;	Mismatches 0;	Indels 2;	Gaps 1;
Qy	1	MVRCIQKLNRTVVAANEICEHFALQPTQACLPICPRDVCVVSBEFLPWSNCKGCGKKLQH	60	
Db	108	MVRCIQKLNRTVVAANEICEHFALQPTQACLPICPRDVCVVSBEFLPWSNCKGCGKKLQH	167	
Qy	61	RTRAVIAPPLFGGLQCNLTETSRACDAPISCPGLGEEYETSLKVGPMKCRLLPHLKEINP	120	
Db	168	RTRAVIAPPLFGGLQCNLTETSRACDAPISCPGLGEEYETSLKVGPMKCRLLPHLKEINP	227	
Qy	121	SGRTVLDPNSDSNERVTFKQSYKAHHHSKSWALEIGYQTRQVSCSTRSDQONAMLSLCLQ	180	
Db	228	SGRTVLDPNSDSNERVTFKQSYKAHHHSKSWALEIGYQTRQVSCSTRSDQONAMLSLCLQ	287	
Qy	181	DSFPLTVQSCIMPXDCETSONSWSPCKTCRSGSLLPGRFSRNRNVKMAIGGKECPE	240	
Db	288	DSFPLTVQSCIMPXDCETSONSWSPCKTCRSGSLLPGRFSRNRNVKMAIGGKECPE	347	
Qy	241	LLEKEACIVEGELLQOCQPRYSWRSEKQVSLLEQQPHHVTPGVCGGGIQTRVY	300	
Db	348	LLEKEACIVEGELLQOCQPRYSWRSEKQVSLLEQQPHHVTPGVCGGGIQTRVY	407	
Qy	301	CAQSVPAALRAKEVRPEKALCVGPAPLPOLNIPCSCTDCTIVSSWAWGLCIHENC	360	
Db	408	CAQSVPAALRAKEVRPEKALCVGPAPLPOLNIPCSCTDCTIVSSWAWGLCIHENC	467	
Qy	361	HEPOGKGFRTORRHVLMESTGPAGHCPHLVESVPCEDPMCYRWLASEGICFPHDCKGL	420	
Db	468	HEPOGKGFRTORRHVLMESTGPAGHCPHLVESVPCEDPMCYRWLASEGICFPHDCKGL	527	
Qy	421	GHRILKAVQNDRGEDVSGSLCPVPPPPERKSCBIPCRMDCVLSEMTWSSCSQCSNKN	480	
Db	528	GHRILKAVQNDRGEDVSGSLCPVPPPPERKSCBIPCRMDCVLSEMTWSSCSQCSNKN	587	
Qy	481	SDGQTSRTTILALAGCGKPCPSQALQHRCLNDHSCMLHWEISPWGPCSEDILTVA	540	
Db	588	SDGQTSRTTILALAGCGKPCPSQALQHRCLNDHSCMLHWEISPWGPCSEDILTVA	647	
Qy	541	LNATIGWGEATCGVGIQTRRVFCVSHVGVMTKRCPDSTRPETVRPCFLPCKDCIYV	600	
Db	648	LNATIGWGEATCGVGIQTRRVFCVSHVGVMTKRCPDSTRPETVRPCFLPCKDCIYV	707	
Qy	601	AFSEWTPCPMCOAGNATVQSRVRIIIOEANGQSCPDTLVEERCEVDVSLCPYRWK	660	
Db	708	AFSEWTPCPMCOAGNATVQSRVRIIIOEANGQSCPDTLVEERCEVDVSLCPYRWK	767	
Qy	661	PQKWSPCILVPESVWQIGTSSACGKGLQTRAVSCISDDNRSAEMMECLQTNMGPLLV	720	
Db	768	PQKWSPCILVPESVWQIGTSSACGKGLQTRAVSCISDDNRSAEMMECLQTNMGPLLV	827	
Qy	721	QECTVPCREDCTTAWSKFTPCSTNCEATKSRRLQTLGKSRKXKCKQDSDLPLVETELC	780	
Db	828	QECTVPCREDCTTAWSKFTPCSTNCEATKSRRLQTLGKSRKXKCKQDSDLPLVETELC	887	
Qy	781	PCDEFISQPYGNWSDCILPEGRREPHGLRVQADSKCEGGLRFAVACSDKNRGPVDPDS	840	
Db	888	PCDEFISQPYGNWSDCILPEGRREPHGLRVQADSKCEGGLRFAVACSDKNRGPVDPDS	947	
Qy	841	FCSSSGYIQEKCVIPCPFDCKLSDWSWGSWSSCGIGVIRSKWLKXKPYNGRCPCKL	900	
Db	948	FCSSSGYIQEKCVIPCPFDCKLSDWSWGSWSSCGIGVIRSKWLKXKPYNGRCPCKL	1007	
Qy	901	DLKN--QVHEAVPCYSECNQYSVWVHWSCKINNELRSIRCGGQTSRKIRCYNVTADGE	958	

Db	1008	DLKQQAQVHEAVPCYSECNQYSVWVHWSCKINNELRSIRCGGQTSRKIRCYNVTADGE	1067	
Qy	959	GAVDNSNLGNODEIPPETQSCSLMCPNECVNWSGLWSKCPQSCDPHTMQRTHLLRPS	1018	
Db	1068	GAVDNSNLGNODEIPPETQSCSLMCPNECVNWSGLWSKCPQSCDPHTMQRTHLLRPS	1127	
Qy	1019	LNSTRCAEDSDVQPCLLNENCFOYQNLTEWSTCOLSENAPCGQGVTRLLSCVCSGDKP	1078	
Db	1128	LNSTRCAEDSDVQPCLLNENCFOYQNLTEWSTCOLSENAPCGQGVTRLLSCVCSGDKP	1187	
Qy	1079	VSMDOCEQHNLEKFORMSIPCLVECVNQCQSGTAWTECSQTCGHGRMSRTRFIIMPT	1138	
Db	1188	VSMDOCEQHNLEKFORMSIPCLVECVNQCQSGTAWTECSQTCGHGRMSRTRFIIMPT	1247	
Qy	1139	QEGSRPCPTELTQKTCPTVPCYSWVLGNWSACKLEGGDCGEGVQIRSLSCMWSHSGSISH	1198	
Db	1248	QEGSRPCPTELTQKTCPTVPCYSWVLGNWSACKLEGGDCGEGVQIRSLSCMWSHSGSISH	1307	
Qy	1199	AAGRVEDALCGEMPFQDSILKQLCSVPCPGDCHLTWSEWSTCELTICIDGRSFETVGRQS	1258	
Db	1308	AAGRVEDALCGEMPFQDSILKQLCSVPCPGDCHLTWSEWSTCELTICIDGRSFETVGRQS	1367	
Qy	1259	RSRFTIIOSFENODSCPOQVLETRPCTGGKCYHYTWKASLWNNNERTVWCORSQGVNVTG	1318	
Db	1368	RSRFTIIOSFENODSCPOQVLETRPCTGGKCYHYTWKASLWNNNERTVWCORSQGVNVTG	1427	
Qy	1319	GCSPOARPAALRQCIIPACRKPFSYCTQGGVCGCEKGYTEIMKSNGLFDYCMKVPGSBDKK	1378	
Db	1428	GCSPOARPAALRQCIIPACRKPFSYCTQGGVCGCEKGYTEIMKSNGLFDYCMKVPGSBDKK	1487	
Qy	1379	ADVKNLSGKNRPVNSKIHDIKFGWSLQPLDPDGRVKIWWYGVSGAGFLIMIFLFTSYLV	1438	
Db	1488	ADVKNLSGKNRPVNSKIHDIKFGWSLQPLDPDGRVKIWWYGVSGAGFLIMIFLFTSYLV	1547	
Qy	1439	CKKXPHOSTPPQKPLTLAYDGLDM	1465	
Db	1548	CKKXPHOSTPPQKPLTLAYDGLDM	1574	
RESULT 2				
AAM41081				
ID	AAM41081 standard; protein; 1551 AA.			
XX	AAM41081;			
AC	AAM41081;			
XX	XX			
DT	22-OCT-2001 (first entry)			
XX	Human polypeptide SEQ ID NO 6012.			
DE	Human; nototropic; immunosuppressant; cytostatic; gene therapy; cancer;			
XX	Peripheral nervous system; neuropathy; central nervous system; CNS;			
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;			
KW	amyotrophic lateral sclerosis; Shy-drager Syndrome; chemotactic;			
KW	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;			
KW	leukaemia.			
XX	Homo sapiens.			
OS	WO200153312-A1.			
PN	26-JUL-2001.			
XX	XX			
PD	26-DEC-2000; 2000WO-US034263.			
XX	XX			
PF	23-DEC-1999; 99US-00471275.			
XX	PR			
XX	21-JAN-2000; 2000US-00488725.			
PR	PR			
XX	25-APR-2000; 2000US-00552317.			
PR	PR			
XX	20-JUN-2000; 2000US-00598042.			
PR	PR			
XX	19-JUL-2000; 2000US-00620312.			
PR	PR			
XX	03-AUG-2000; 2000US-00653450.			
PR	PR			
XX	14-SEP-2000; 2000US-00662191.			
PR	PR			
XX	19-OCT-2000; 2000US-00693036.			

29-NOV-2000; 2000US-00727344.  
(HYSE-) HYSEQ INC.  
Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;  
Zhou P, Goodrich R, Drmanac RT;  
WPI; 2001-442253/47.  
N-PSDB; AA160237.  
Novel nucleic acids and polypeptides, useful for treating disorders such  
as central nervous system injuries.  
Example 2; SEQ ID NO 6012; 10078pp; English.  
The invention relates to human nucleic acids (AA157798-AA161369) and the  
encoded polypeptides (AA158442-AA162213) with nootropic,  
immunosuppressant and cytostatic activity. The polynucleotides are useful  
in gene therapy. A composition containing a polypeptide or polynucleotide  
of the invention may be used to treat diseases of the peripheral nervous  
system, such as peripheral nervous injuries, peripheral neuropathy and  
localised neuropathies and central nervous system diseases, such as  
Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
utilisation of the activities such as: Immune system suppression,  
Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
assays for receptor activity, arthritis and inflammation, leukaemias and  
C.N.S disorders. Note: The sequence data for this patent did not form  
part of the printed specification

Sequence 1551 AA;  
Query Match 52.1%; Score 4296; DB 4; Length 1551;  
Best Local Similarity 50.7%; Pred. No. 2.8e-246;  
Matches 766; Conservative 238; Mismatches 449; Indels 59; Gaps 20;

2 VRCIQKNTTVANETCEHAFALQPTTEQACLPICRDRCVSEFLPWSNCKGCGKKLQHR 61  
52 IACIQKNDIPAEADIIICEYFEPEKPLEEQACLPICQDQCIIVSEFSAWSECKTKCGSLQHR 111  
62 TRAVIAPPLFGGLQCNLTESACDAPISCPGEEYFTSLXVGPWSKRLPLHKEINPS 121  
112 TRHVAPPOFGSGCNPNTLTFVQVQ---SSPEAEELRYSLHVGFWSTCSMPSHQVQQA 169  
122 GRT-----VLDNFSDSNERNVTFK-HQSYKAHHHSKSWAIEIGYQTRQVSCTR 167  
169 RRRGKNKEREKDRSKGVKD--PEARELIKKNRNQRQENKYWDIIGYQTRVWMCIN 226  
168 SDQGNAMLICLQDSPPLTVQSCIMPQDCTQWSNWSPCSKTCBSGILLPGFRSRNV 227  
227 KTGKAADLSFCQOEKLPMTFQSCVIITKEQVSEWSEWSPCKTCHDMVSPAGTRVTRTI 286  
228 XMAIGGKECEPELLEKEACIIVEGELLQCPRYSWRTSEWKECVSLLEQQDPHHVYG 287  
287 RQFFIGSEKEFEFEKEPLSQDGVWPCATYGNATTETECRVDPLLSQDKERGNQT 346  
288 PVCGGIQTRVYCAQS-----VPAALRAKEVRSPVEKALCVGPAPLPSQLCNTPCSTD 343  
347 ALCGGGIOTREYVCYQVQANENLLSLSLTHNKEASKFMDLKLCTGPIPTNTQLCHIPCPE 406  
344 CTVSSWSAMGLCITHENCHPEQGGKGFRTQRHVLMESTGPA---GHCPLHVESVCEPDM 400  
407 CEVSPHANGPTCYENCNDQKQKGFLEKRIITNEPTGGSGVTGNCPHLLLEAIFCEEPA 466  
401 CYRWLASE-GICFPDHGK-CGLGHRILKAVCONDRGEDVSGSLC-----PVPPPERKSC 453  
467 CYDWKAVRLGDCPEPDNGKEGPGTQVQEVVNCINSDEGEVDRLCRDAIFPIP-----VAC 521  
454 ETPCRMDCVLTSEWTESSCSGSCNKNKSGKQTRRTILALAG-EGGKPCPPSQALQHR 512  
522 DAPCPKDCVLTSTWSTSSCSHTCSGKTTGKQIRAKSILAYAGEGGIRCPNSALQVR 581

QY 513 LCNDSHSCMLHWTSPWGPCSEDTLVTALNATIGWGEATCGVIGIOTRRVFCVKSHVGV 572  
Db 582 SCNEHPTCVVHWOTGPGWQCIEDTSVSSFNITTTWNGEASCSVMQTRKVICRVNVGV 641  
QY 573 MTKRCPDSTPEVTRPCFLPKCKKDCITVAFSEWTPCPRMCOAGNATV-KQSRVRIIIEA 631  
Db 642 GPKKCPESLAPETVRPCLLPCKKDCITVTPYSDWTSPPSSCKEGSSIRKQSRHVIQLP 701  
QY 632 ANGQBCPDTLIYERBCEDESLSLCPVVRWPKQKWSPCILIPESVWQITGSSEACGKLQT 691  
Db 702 ANGRDCTDLYEKEACEAPQACQSYRWTWKHRCQLVPSVQODSPGAQEGCGPQA 761  
QY 692 RAVSCISDDNRSAAEMMECKQTMGMPLLVQECTVPCKEDCTFTAMSKFTPCSTNCEATKS 751  
Db 762 RAITCRKQDQAGIHECLQYAGVPALTAQCOIPQDDCOLTSWKSFCSCNGCCAVRT 821  
QY 752 RRQLTKSKKKBKQSDILYPLVETELCPDFFISQPYGNWSDCILPEGRRPHGLRV 811  
Db 822 RKTLVGSKKKECKNSHLYPLIETQYCPDKYNAQPVGNWSDCILPEGVLLGMKV 881  
QY 812 QADSKEGGLRFRVACSDKNGRPVDPSPCSSGYIOEKVIPCPDCKLSDWSSWGSC 871  
Db 882 QGDIKEGCGGIRQAWACYDQNGRLVETSRNSHGHIIEACIIPCFSDCKLSWNSRC 941  
QY 872 SSSCGIGVIRSKWLKEKPYNGRPPCKLDLKN--QVHEAVPCYSECNQYSWVVEHWS 929  
Db 942 SKSGSGVKVRSKWLREKPYNGRPPCKLDHVNQAQVYVWVPSCHSDCNOYLWTEPWSIC 1001  
QY 930 KINNELSLRCPGGTQSRKIRCV-NTADGEGGAVDSNLQNDIEIPPETOSCSLMCNECV 988  
Db 1002 KVTFFVNRNRENGEGVQTRKVRQMONTADGSEHVEDYLCDEEMPLGSRVKLPCPEDCV 1061  
QY 989 MSEWGLWSKCPQSCDPHTMORTREHLIR-PSLSNRITCAEDSQVQPCLLNENCFQFYNLT 1047  
Db 1062 ISENGPWTCVLPQCNQSSFPQRSADPIQPADGERSCPNAVEKEPCNLKNKYHYDNTV 1121  
QY 1048 EWSTCQLSENAQCQGVTRRLLSVCVSDGKPVSDMOGEOHNLKPKORMSIPCLVECVNC 1107  
Db 1122 DWSTCQLSEKAVCGNGIKRMLDCVRSDGSKSVDLKYCEALGLEKNQMNTSCMVECPVNC 1181  
QY 1108 QLSGWTATWCSQTCGHGGRMSRTRFIIMPTQCEGRPCPTLTQETCTPVTPCYSVVLGN 1167  
Db 1182 QLSDWSPWSECSQTCGLTGKMIERRTVTPFQDGRPCPSLMDQSPKPCVRYHQYG 1241  
QY 1168 WSACKLEGPCGEGVQIRLSLSCMVHSGSISHAAGRYEDALCGEMPF-----QDSILKQLC 1222  
Db 1242 WSPCQVQEAQCGEGTRTNISCVVSDGSDDDFSKVVDDEEFCADIELIIDGNKNWLEESC 1301  
QY 1223 SVPCPGDCHLTSEWSEWSTCELTCTIDGRSPETVGRQSRRTFIIOSEFENODSCPOQVLETR 1282  
Db 1302 SQPCPGCYLKMSSWSLQCLTCVNGEDLGFQIGVRSRVIIQELENOHLCEQMLETK 1361  
QY 1283 PCTGGKCYHTWKAASLWNNNRTWVQCRSDGVNVTGCSFQAPAAIROCIIPACRPFYS 1342  
Db 1362 SCYDGCQYEVKMAWASAKSGSRVTCQSDGINVTGCLVMSQPDADRSNPPCSQPHSY 1421  
QY 1343 CTQGVGCGCKGYTEIMKNGFLDYC-----MKVPSSEDKADYKLSGKRRPVNSKIHD 1397  
Db 1422 CSETKTCHCEGTEVMSNSSTLEQCLTIPVVVLPMTMEDKRGDKVT-SRAVHTPTQSSNP 1480  
QY 1398 IFKG--WSLOPLPDGRVKLWYVGGGAFILIFLIFTSYLVCKKP-KPHQSTPTPOQKP 1454  
Db 1481 AGRGRTWFLQFPDGLKLTWVGVGAAGAVLLIFIVSMYILACKKPKPQRQNNLKP 1540  
QY 1455 LTLAYDGLDM 1465  
Db 1541 LTLAYDGLADM 1551

RESULT 3  
AAB20155  
ID AAB20155 standard; protein; 1598 AA.







573 MTKCPDSTRPETVRPCFLPKCKDCIVTAFSEWTPCFRMCQAGNATV-QOSRYRIIIOEA 631  
 679 GPXKCPESLPETVRPCFLPKCKDCIVTAFSEWTPCFRMCQAGNATV-QOSRYRIIIOEA 738  
 632 ANGQCEPDLTYEERECEDVSLCPVYRWKQKNSPCILIPESVWQITGSSEACGKLOT 691  
 739 ANGRCDDTPLYEKEACEAPQACQSVRWKTHKRRCOLVPWSVQDQSPGACGCGPQRA 798  
 692 RAVSCISDDNRSAEMMECLLQKTNGMPLLVQECTVPCREDCTFTAWSKFTPCSTNCEATKS 751  
 799 RAITCRKQDGGQAGIHECLQYAGVPALTAQCOIPQDDCQLTWSKFSKSCNGDCGAVRT 858  
 752 RRRLQTKSRKKEKQSDILYPLVETELCDCEFIQPYGNWSDCILPGRREPHRGLRV 811  
 859 RKETLVGKSKKXKCKSHLYPIETQYQCDPCDYNAPQVGNWSDCILPGRKVEVLGMKV 918  
 812 QADSKEGEGIRFRAVACSDKNGRPVDPSCSSGYIQEKVCIPCFDFCKLSDWSWGSC 871  
 919 QGDIKECGQGVRYQAMACYDONGRLVETSRCSHGYIEEACIIPCPSDCKLSEMSNKSRC 978  
 872 SSSCGIGVIRISKWKKEKPNGRPCPKLDLKN-QVHEAVPCYSEGNQYSVWVHWSSC 929  
 979 SKSCGSGVKRSKWLREKPNGRPCPKLDLKN-QVHEAVPCYSEGNQYSVWVHWSSC 1038  
 930 KINNELSLRCGGGTQSRKIRCV-NTADGGGAVDSNLNODIPEPTOSCSLMCPNECV 988  
 1039 KVTFFVNRNCGEGVQTRKVRCHQNTADGSEHVEDYLCDEEYPLGSRVCKLPCPEDCV 1098  
 989 MSEWGLNSKQPSQCDPTMQRTHLLR-PSLNSRTCAEDSQVQPCLLNENCSFQFYNT 1047  
 1099 ISEWGPMTQCVLPNCQSSFRQSRADPIRQPADEGRSCPNNAVEKPECNLKNVCHYDYNV 1158  
 1048 EWSFCQSENAQCGQVTRLLSCVSDGKPSVSDQCEQHNLEKQFMSIPCLAVECVNVC 1107  
 1159 DWSTCQSEKAVCGNGIKRMLDCVSDGKSVLDKCEALGLSKNWMQNTSCMVECPVNC 1218  
 1108 QLSGWTATWCSQTCGRGRMSRTRFIIMPTQEGRCPTELTQEKTCPTVTPCYSVWLG 1167  
 1219 QLSWSPWSECSQTCGLTKMIRRTVTQPFQDGRPCPSLMDQSKPCPKVPCYRQYQG 1278  
 1168 WSACKLEGCGEGVQIRSLSCVHSGSISHAAGRYEDALCGEMPF-----QDSILKQLC 1222  
 1279 WSPCQVQEQAGCGEGTRTRNSCVVSDGSADDFSKVVDDEFCADIELIIDGNKNWLEESC 1338  
 1223 SVPCFGCHLTWSEWSTCBELTICDGRSFETVGRQSRRTFIIQSPENDQSCPQVLETR 1282  
 1339 SQPCPGCYLKDWSSWSLQCLTCVNGEDLFGGIQVRSRPVLIQELNQHLCPEQMLETK 1398  
 1283 PCTGCKCYHTWKASLWNNERTVWCORSQGVNVTGCSQAPAPAIROCIACRKPFSY 1342  
 1399 SCYDQCYEYKWMASAKSGSRTVWCORSQGVNVTGCSQAPAPAIROCIACRKPFSY 1458  
 1343 CTQGGVCGCEKGYTEIMKSNGLDYC-----MKVPGSEDKKADVKNLSGKRRPVNSKIH 1397  
 1459 CSETKTCHCEGYTEVMSNSTLEQCLTLPVVLPTMEDKRGVKT-SRAVHTQPSNP 1517  
 1398 IFKG--WSLQPLDPGRVKVWVGVSGAFILMIFLFTSVLVCKXP-KPHQSTPPQOKP 1454  
 1518 AGRGRTWFLQFGPDGRKLTWTVGAAGAEVLLIFVSMIYLACKPKPKPQRQNNRLXP 1577  
 1455 LTLAYDGDLD 1465  
 1578 LTLAYDGDADM 1588

RESULT 5

ABR58462

ID ABR58462 standard; protein; 1588 AA.

AC ABR58462;

XX ABR58462;

XX 07-JUL-2003 (first entry)

DT 07-JUL-2003 (first entry)

XX 07-JUL-2003 (first entry)

XX 07-JUL-2003 (first entry)

Human NOV47m.

Human; NOV; antidiabetic; anorectic; antibacterial; virucide;  
 immunomodulator; cytosolic; nootropic; neuroprotective; dyslipidaemia;  
 aniparkinsonian; antilipemic; gene therapy; metabolic disorder;  
 diabetes; obesity; infection; cachexia; cancer; Parkinson's disease;  
 neurodegenerative disorder; Alzheimer's disease; immune disorder;  
 haematopoietic disorder.

Homo sapiens.

WO2003029423-A2.

10-APR-2003.

02-OCT-2002; 2002WO-US031358.

02-OCT-2001; 2001US-0326483P.

05-OCT-2001; 2001US-0327342P.

09-OCT-2001; 2001US-0327917P.

09-OCT-2001; 2001US-0328029P.

09-OCT-2001; 2001US-0328044P.

09-OCT-2001; 2001US-0328056P.

12-OCT-2001; 2001US-0328849P.

15-OCT-2001; 2001US-0329414P.

17-OCT-2001; 2001US-0330142P.

22-OCT-2001; 2001US-0341058P.

24-OCT-2001; 2001US-0339286P.

24-OCT-2001; 2001US-0343629P.

29-OCT-2001; 2001US-0349575P.

01-NOV-2001; 2001US-0346357P.

12-APR-2002; 2002US-0371972P.

12-APR-2002; 2002US-0371980P.

17-APR-2002; 2002US-0373281P.

19-APR-2002; 2002US-0373805P.

23-APR-2002; 2002US-0374738P.

16-MAY-2002; 2002US-0381101P.

29-MAY-2002; 2002US-0381635P.

01-OCT-2002; 2002US-0383830P.

01-OCT-2002; 2002US-00262839.

(CURA-) CURAGEN CORP.

Alsbrook JP, Anderson DW, Boldog FL, Burgess CE, Catterton E;  
 Edinger SR, Ellerman K, Gerlach VL, Gorman L, Guo X, Ji W;  
 Kekuda R, Leach MD, Li L, Miller CE, Patturajan M, Rieger DK;  
 Rothenberg ME, Shinkens RA, Smithson G, Spytek KA, Taupier RJ;  
 Vernet CM, Voss EZ, Zethusen BD, Zhong M;

WPI; 2003-381625/36.

N-PSDB; ACC72174.

NOVX polypeptides and nucleic acids useful for diagnosing, preventing or  
 treating NOVX-associated disorders, e.g. diabetes, obesity, cancer or  
 dyslipidaemia, and in chromosome mapping, tissue typing or  
 pharmacogenomics.

Claim 1; Page 267; 487pp; English.

The present invention relates to novel human NOV proteins and their  
 coding sequences (ACC72075-ACC72181 and ABR58363-ABR58469). The NOV  
 proteins are useful in manufacturing a medicament for treating a syndrome  
 associated with a human disease. The NOV proteins and coding sequences  
 may be used to diagnose, treat or prevent metabolic disorders such as  
 diabetes or obesity, infections, cachexia, cancer, neurodegenerative  
 disorders such as Alzheimer's disease or Parkinson's disease, immune  
 disorders, haematopoietic disorders and various dyslipidaemias

Sequence 1588 AA;

Query Match 52.1%; Score 4296; DB 6; Length 1588;

Best Local Similarity 50.7%; Pred. No. 2.9e-246;

Matches 766; Conservative 238; Mismatches 449; Indels 58; Gaps 20;

1048 EWSTCOLSENAPCGQVTRILLSCVSCDGPYSMDQCEQHNLEKQPMRSIPCLVECVVNC 1107  
1159 DWSTCOLSEKAVCGNGIKTRMLDCVSDGSKVDLKYCEALGLEKNQWMTSCWVECPVNC 1218  
1108 QLSGWTAWTECSOTCGHGRMSRTRIIIMPTQGEGRPCPTTELTOEKTCPVTTCYSWVLGN 1167  
1219 QLSDWSPWSECSQTCGLTKMIRRTVTQTFQDGRPCPSLMDQSKPCVPKPCYRMYGQ 1278  
1168 WSACKLEGCGGCGVOIRSLSCMVHSGSISHAAGRVEDALCGEMPF-----ODSIILKQLC 1222  
1279 WSPCVQEAQCGEGRTRNISCVVSDGSDADDPSKVVDDEFCADIELIIDGNKNWLEESC 1338  
1223 SVPCPDGCHLTWEMSESTCELTICIDGRSPETVGRQSRRTFIIQSPENQDSCPQVLETR 1282  
1339 SQPCPDGCVLKDWSHSLCQLTCVNGEDLFGGQIVRSRPVITQELNHLCPQEMLETK 1398  
1283 PCTGGKCYHYTWKASLWNNERTVWCQSDGVNVTGCGSPQAPPAIROCIIPACRPFESY 1342  
1399 SCVDGQCYEYKMASAKGSSRTVWCQSDGINVTGCLVMSQPDADRSCNPPCSQPHSY 1458  
1343 CTQGVGCGCKGYTEINKNGFLDYC-----MKVPGSEDKADVKNLGKNRPVNSKIH 1397  
1459 CSETKTCHCEGTEVWSSNSTLEQCTLIPVVVLPTMEDKGDVKT-SRAVHPTQSSNP 1517  
1398 IFKG--WSLOPLDPDGRVKIWWYGVSGGAFIMFLIFTSYLVCKP-KPKQSTPPQOKP 1454  
1518 AGRGRTWLQFPFGPDGRLLKTMVYGAAGAFVILIFIVSMYILACKPKPKPQRQNNRLXP 1577  
1455 LTLAYDGDLD 1465  
1578 LTLAYDGDADM 1588

RESULT 6  
AAE23979  
ID AAE23979 standard; protein; 1624 AA.  
XX AAE23979;  
XX  
XX 23-SEP-2002 (first entry)  
XX Human LP217 secreted protein.  
XX  
XX Human; secreted protein; atherosclerosis; Alzheimer's disease; LP217;  
XX diabetic retinopathy; severe combined immunodeficiency; pancreatitis;  
XX rheumatoid arthritis; colorectal adenoma; haemolytic anaemia; cancer;  
XX reperfusion injury; arteriosclerosis; wound healing; transgenic animal;  
XX gene therapy; neoplasm; transgenic; psoriasis; ischaemia; carcinoma;  
XX chromosome 7p21-p22.  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
XX Peptide 1..15  
XX Protein /label= signal\_peptide  
XX 16..1624  
XX /note= "Mature human LP217 secreted protein"  
XX  
XX W0200226801-A2.  
XX  
XX 04-APR-2002.  
XX  
XX 14-SEP-2001; 2001WO-US026026.  
XX  
XX 26-SEP-2000; 2000US-0236088P.  
XX  
XX (ELIL ) LILLY & CO ELI.  
XX  
XX Su EW, Wang H;  
XX  
XX WPI; 2002-471259/50.  
XX N-PSDB; AAD38694.  
XX

2 VRCIQKLNRTVVANBICEHFALOPTEQACILPCPRDCVVSSEFLPWSNCSKGGCKLQHR 61  
89 IACIQKDKDPAEDIIICEYFEKPLEQACILPCQODCIVSEFSANSECSKTCGSLQHR 148  
62 TRAVTAPPLFGGLQCPNLTESACDAPISCPGEBEYFSLKVGWSPSKRILPHLKEINPS 121  
149 TRHVAPPOFGSGCPNLTEFQVCQ---SSPCEAEELRYSLHVGFWSTCSPHSPQVRAQ 205  
122 GRT-----VLDNFSDSNERTVEFK-HOSYKAHHHSKSWAIBGYCTROVSCTR 167  
206 RRRGNKREKDRSGKVD--PEARLKKGNRNRQENKYWDIQIGTREVWMCIN 263  
168 SQQNAMLSLCI-ODSFPLTVQSCINPKOCETSQMSWSPSCSKTCRSGSLLPFRSRSNV 227  
264 KTGKAADLSFCQOEKLPMTFQSCVITKEQVSEWSEWSPCSKTDHMSVPASTRVTRTI 323  
228 KMAJGGGKECEPELLEKEACIVGELLQCPRYSWRTSEWKECVSLLEOODPHHVTG 287  
324 RQFP-IGSEKCEPEFEKEPCLSGQGVVPCATYGRWTEWTECRVDPLLSQDCKRGNQT 383  
289 PYCGGGIQTREYVCAQS-----VPAAALRAKEVSRPVEKALCVGPAPLPSOLCNIPCS 343  
384 ALCGGGIQTREYVYVQANENLLSQLSTHKNKSAKPMDLKLCCTGPI-PNTTQLCHIPCPT 443  
344 CIVSSNSAGLCIHENCHPEQCKGFRTRORHVLVESTGPA---CHCPHVESVPCEDPM 400  
444 CEVSPMSAWGPTCYENCNDQOGKGFGRKRRTITNEPTGGGVTGNCPHLEAIPCEPA 503  
401 CYRWLASE-GICFPDHGK-CGIGHRIKAVCONDRGEDVSGSLC-----PVPPPERKSC 453  
504 CYDWKAVRLGDCPEPNGKECGPTQVQEVVCLNSDGEVDRLCHDAIFPFP-----VAC 558  
454 ETPCRMDCVLSEWTEWSSQSCSNKNSDGKQTRSRKTILALAG-BGGKPCPSPQALQHR 512  
559 DAPCPKDCVLSTWSTWSSCSHTCSGKTGKQIRARSILAYAGEGGIRCPNSSALQVR 618  
513 LCNDRSCMLHWETSPWGPCSDDTLVLANATIGNWGATCGVGTQTRRVFCVSHVQGV 572  
619 SNEHPCVYHHQTGPWGQCIEDTSVSSFNITTTWNGEASCSVGMQTRKVICRVNVQGV 678  
573 MTKRCPDSTRPRTVPCFLPCKKDCIVTAFSWTPCPRMCOAGNATV-KQRYRIIIOEA 631  
679 GPKKCPESLRPRTVPCLLPCKKDCIVTFSWTSKPSCKEGDSSIRKQSRHVRVILQ 738  
632 ANGGOCPPLVEERECEDVSLCPVYRWKPKQWSPCILLVPESVMQGITGSSEACKGLQT 691  
739 ANGRDCTDPLVEEKACEAPQACQSYRWKTHKWRRCQLVPMVSQDPSGAQEGGPGRQA 798  
692 RAVSCISDDNRSAEMMECLKQTNMGPLLVQECTVPCREDCTFTAMSKFTPCSTNCEATKS 751  
799 RAITCRKQDGGQAGIHECLQYAGVPALTQACQIFCQDDCQLTSWKFSSCNGDCGAVRT 858  
752 RRQLTGKSRKEKQODSLYPLVETELCPDEBFTSQYGNWSDCILPEGRPREHRGLRV 811  
859 RXRTLVGKSKKEKCKNSHLYPLIETQYCPCKDYNAQVPVGNWSDCILPEGKVEVLLGKV 918  
812 QADSKECEGLRFAVACSDKNRPVDRPSFCSSGVIOEKVIPCPCFCKLSDWSWGS 871  
919 QGDIKECGGYRYQAWACYDQNGRLVETSRCHSHGYIEACIIPCPSCKLSEWNSWRC 978  
872 SSSCGIGVIRSKWLKEKPYNGRCPKLDLKN--QVHEAVPCYSECNQYKSWVEHWSSC 929  
979 SXSCSGSVKRSKWLREKPYNGRCPKLDHVNQAQVVEVPCHSDCNQYLWVTEPWSIC 1038  
930 KNNELSLRCCGGTOSRKIRCV-NTADGEGAVDSNLQNDIEIPETQSCSLMCPNECV 988  
1039 KVTFTVNRNCEGEGVTRKVRQMTADGPSHEHSDYLCDEEMPISGRVCKLPCPEDCV 1098  
989 MSEWGLWSKCPQSDPHMTQRTREHLR--PSLNSRTCAEDSQVPCLLNENCFQYQNL 1047  
1099 ISEWGPWTQVLPNCQSSPQRSADPIRQPADEGRSCPNAVEKPCNLNKNCHYDYNVT 1158

Novel proteins and polynucleotides of secreted proteins useful for treating various diseases e.g. rheumatoid arthritis, cancer, psoriasis, diabetic retinopathy, arteriosclerosis, ischemia or reperfusion injury.

Claim 8; Page 117-123; 145pp; English.

The invention relates to human secreted polypeptides designated LP095, LP191, LP217, LP220, LP221, LP222, LP229, LP237 or LP238 and nucleic acid molecules encoding such polypeptides. Novel secreted proteins of the invention are used for treating diseases such as arteriosclerosis, Alzheimer's disease, diabetic retinopathy, psoriasis, pancreatitis, arteriosclerosis, rheumatoid arthritis, colorectal adenomas, severe combined immunodeficiency, ischaemia, carcinomas, haemolytic anaemia, reperfusion injury, neoplasms and cancer especially liver cancer. They are also used for wound healing. Polynucleotides of the invention can be used to generate transgenic animals or knock out animals, which in turn, are useful in the development and screening of therapeutically useful reagents for use in the treatment of diseases associated with LP polypeptide associated activity. They are also used in gene therapy. The present sequence is human LP217 secreted protein. LP217 gene is located on chromosome 7p21-p22

Sequence 1624 AA;

Query Match 52.1%; Score 4296; DB 5; Length 1624;  
Best Local Similarity 50.7%; Pred. No. 3e-246;  
Matches 766; Conservative 238; Mismatches 449; Indels 58; Gaps 20;

2 VRTQKLNRTVANEICHEPALPPTFEQAQLPCPRDCVVFSPNSKCGKGLQHR 61  
125 IACIQKDKIPAEIDIEYEPKPLLEQAQLPCQODCIVSEPSAWSECKTGGSLQHR 184  
62 TRAVIAPPLFGGLQCPNLTESRACDAPISCPLEEEYTSIKVPSKCLPHLKEINPS 121  
185 TRHVAVAPPFGGSGCPNLTBFQVCQ---SFCBAELRLVSHVGPWSTCMPSHRSRQVQA 241  
122 GRT-----VLDNFSDNSNRVTFK-HQSVKAAHSHKSWAIEIGYQTRQVSCRT 167  
242 RRRGNKERKDRSGKYKD--PEARELLIKKRNRRQNRQENKYWDIQGYQTRVWCIN 299  
168 SDGNAMLSLQDSFPLTVQSCIMPMDSTQNSWSPCKTGRSGSLLPGRSRNV 227  
300 KTGKAADLSFCQBEKLPMTQSCVITKEQVSEWSEWSPCKTCHDWSPAGTRVTRTI 359  
228 KMAIGGKECPLELEKACIVGELLQCCPRYSWRTESEKQVSLLEQDPHWHVTG 287  
360 RQFFIGSEKEPEPEKEPCLQSQDGVVPCATVGRWTEWTECRVDPPLSQDXRGNQT 419  
288 PVOGGGIQTRVYCAQS-----VPAALRAKEVSRPVEKALCVGPAPLPSPQLCNIPSTD 343  
420 ALCGGGIQTRVYCVQANELLSQLSTHKNKEASKPMDLXLCTGP-PNTTQLCHIPCPT 479  
344 CIVSSWSAGLCIENCHPEQKGFTRORHVLMESTGPA---GHCPLHVESVPCEDPM 400  
480 CEVSPWAGWGCCTYENCNDQGGKGFKLRRRIINEPTGGSGVTGNCPLLEAIPCEEPA 539  
401 CYRWLASE-GICFPDHGK-CGLGHRILKAVCONDRGEDVSGSLC-----PVPPPPKSC 453  
540 CYDNKAVRLGDCPEFNGKECGPGTVQVQVWCINSDEGEVDGRLCRDAIFPIP-----VAC 594  
454 EIPCRMDVLSEWTEWSSCSOSCKNSDKQTRSRITILALAG-EGKPCPPQALQHR 512  
595 DAPCPKCVLSTWSTWSSCSHTSGKTKTEGQIRARSILAYAGEEGIRCPNPSALQEV 654  
513 LCNHSCNQLHWETSPWGPCSEDTLVLTALNATIGNCEATCGVIGTRRVFCVKSHVGV 572  
655 SCNEHPCVTHWQCPGQCIEDTSVSEFNTTTTNGEASCSVGMQTRKVICRVNVGV 714  
573 MTKGCPDSTRBTRVPCLPKCKDCIVTARESWTPCPRMQAGNATV-KOSRVRIIIOEA 631  
715 GPXKCPESLRBTRVPCLLPKCKDCIVTARESWTPCPRMQAGNATV-KOSRVRIIIOEA 774  
632 ANGQCEPDLTYERECECDLSLCPVRWKPKQKSPCILVPESVWQGITGSACGKLQT 691

## RESULT 7

AAM39295

ID AAM39295 standard; protein, 1588 AA.

XX

AC AAM39295;

XX

DT 22-OCT-2001 (first entry)

XX

DE Human polypeptide SEQ ID NO 2440.

XX

Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
peripheral nervous system; neuropathy; central nervous system; CNS;  
Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;

amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
chemokinetin; thrombolytic; drug screening; arthritis; inflammation;  
leukaemia.

Homo sapiens.

WO200153312-A1.

26-JUL-2001.

26-DEC-2000; 2000WO-US034263.

23-DEC-1999; 99US-00471275.

21-JAN-2000; 2000US-00488725.

25-APR-2000; 2000US-00552317.

20-JUN-2000; 2000US-00598042.

19-JUL-2000; 2000US-00620312.

03-AUG-2000; 2000US-00653450.

14-SEP-2000; 2000US-00662191.

19-OCT-2000; 2000US-00693036.

29-NOV-2000; 2000US-00727344.

(HYSB-) HYSBO INC.

Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;

Zhou P, Goodrich R, Drmanac RT;

WPI; 2001-442253/47.

N-PSDB; AA158451.

Novel nucleic acids and polypeptides, useful for treating disorders such

as central nervous system injuries.

Example 4; SEQ ID NO 2440; 10078pp; English.

The invention relates to human nucleic acids (AA157798-AA161369) and the  
encoded polypeptides (AA138442-AA142213) with nootropic, and the  
immunosuppressant and cytosolic activity. The polynucleotides are useful  
in gene therapy. A composition containing a polypeptide or polynucleotide  
of the invention may be used to treat diseases of the peripheral nervous  
system, such as peripheral neuropathies and central nervous system diseases, such as  
Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
utilization of the activities such as: immune system suppression,  
Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
assays for receptor activity, arthritis and inflammation, leukaemias and  
C.N.S disorders. Note: The sequence data for this patent did not form  
part of the printed specification

Sequence 1588 AA;

Query Match 52.1%; Score 4292; DB 4; Length 1588;

Best Local Similarity 50.6%; Pred. No. 5e-246;

Matches 765; Conservative 239; Mismatches 449; Indels 58; Gaps 20;

2 VACIOKNTVAVNEICEFALQPPTEQAQCLIPCRDCCVVSFPLWSNCKGCGKLIQHR 61

89 IACIQKXDIAPEDIIICEFYEPKPLLEQACLIPOQDCIVSFSAWSBCKTCGSGLQHR 148

62 TRAVIAPPLFGGLQCPNLTESRACAPISCLGEBEYTFSLXVGPWSKRLPHLKEINFS 121

149 TRHVAPPFGSGGCPNLTEFVQCQ---SSPCEAEELRYSLHVGWSTCSMPHSQVQA 205

122 GRT-----VLDFNSDSNERVTFK-HQSYKAHHHKSVAIIGIQTROVSCTR 167

206 RRGKKNKEREKORSKGVKD--PEARLKKKNRNQRQENKQYWDIQIGYTREVMCIN 263

168 SDGONAMLSLCLQDSPLTVQSCIMPKDCSETQSSWSWSPCKSTCRSGSLLPGRFSRNV 227

264 KTGKAADLSFCQEKLPMTFQSCVITKEQVSEWSWSPCKSTCHDMVSPAGTRVTRTI 323

QY 228 KMAIGGKCEPDELLEKEACIVGELLQOCFPRYSWRTSEWKEQVSLILLQODPHHMTG 287  
DB 324 RQFFIGSEKCEPFEKEPCLSQDGVVPCATYGRITTEWTRTECRVDPLLSQDQRRGNQT 383  
QY 288 FVCGGG:QIREVYCAQS---VFAAALRAKESRVPVEKALCVGPAPLPSPQLCNIPGSTD 343  
DB 384 ALCGGG:QIREVYCVQANELLSQLSTHKNKEASKPMDLKLCTGPIPTNTQLCHIPCPE 443  
QY 344 CIVSSNSAWG:CIHENCHBPQGGKGFTRORHVLMTSTGPA---GHCPHLVESVPCEPDM 400  
DB 444 CEVSPNSAWGPECTYENCNDQGGKPKLRKRITNEPTGGSGVTGNCPHLEAIPCEPA 503  
QY 401 CYRWLASE-GICFPDGHK-CGLGHRILKAYCQNDRGEDVSGSLC----PVPPPPERKSC 453  
DB 504 CYDKAWYRLGDCPEPDNGKECGPTQVQEVVVCINSDEEDVRLQCRDAIFIP-----VAC 558  
QY 454 EIPCRMDCVLSEWTESSQSCSNKNSDKQTRSRITILALAG-EGKPCPPPSQALQEH 512  
DB 559 DAPCPKDCVLSWTWSSCSHTCSGKTTEGQIRARSILAYAGEEGIRCPNSSALQEV 618  
QY 513 LCNDSHSCMLHWETSPMGPCSEDTLVATLNATIGWGEATCGVGIOTRRVFCVKSHVGV 572  
DB 619 SCHEHPTVTHGTGPMGQCIEDTSVSSFNITTTWAGEASCVSVMQTRKVICURVNVGV 678  
QY 573 MTKRCPDSTRTPTVRPCLPKKDCIIVTAFSEWTPCFRMCQAGNATV-KQSRYRIIIEA 631  
DB 679 GPXKCPESLRPTVRPCLLPCKKDCIIVTYSWDTSCPSSCKEGDSIRKQSRHRVILP 738  
QY 632 ANGQECPTLYBERSECEVSLCPVVRWKPQKSPCLILVPESVWQITGSSEACGGLQT 691  
DB 739 ANGRDCTDPLYEKEACEAPQACQSTRWTKHRRRCQLVPMVQVQSDPSGAQEGCGPQ 798  
QY 692 RAVSCISDDNRSAAEMMECLQTMGPLLVQECVTPCREDCFTFAMSKFTPCSTNCEATKS 751  
DB 799 RAITCRQDGGQAGIHECLQYAGVPALTAQACIPQDDCQLTSMKSFSSCNGDCGAVR 858  
QY 752 RRQLTKSRKKEKQDSDLYPLVETELCPDEFISQPVGNWSDCILPEGRRPHGLRV 811  
DB 859 RKTLVGSKSKKKCKNSHLYPLIETQYCPDKYNAQPVGNWSDCILPEGKRVLLGMKV 918  
QY 812 QABSKEGEGRLRPAVACSDKNGRPVDPSPCSSSGYQEKVPCPPFCDKLSWWSWGC 871  
DB 919 QGDIKEGCGQYRVQMACYDONGELVETSECNHSHVIEACIIPCSDCKLSWWSWRC 978  
QY 872 SSSCGIGVIRTRSKWLKEKPYNGRCPCKLULKN--QVHEAVPCYSCNQYVWVHSSC 929  
DB 979 SKCGSGVKVRSKWLREKPYNGRCPCKLHVNAQAVYEVVPCSDCNQYLVWTEPWSIC 1038  
QY 930 KINNELSLRCGGGTOSRKIRCV-NTADGEGGAVDSNLGNODEIPPETOSCSLMCPNECV 988  
DB 1039 KTVFVNRNCGEGVQTRKVRCMQNTADGPSEVEDYLCDPEMPLGSRVCKLPCPEDCV 1098  
QY 989 MSBWGLWSKCPQSCDPHTMORTRHLLR-PSLNSRTCAEDSQVQPCLLNENCFQOYNLT 1047  
DB 1099 ISEWGPWTQVLPFCNQSSFRQSRADPIQPADBGRSCPNAVEKEPCNLNKNKCYHYDNYT 1158  
QY 1048 EWSNCSLSEAPCGQVTRTELLSCVSDCKPFYSDMOCEQHNLEKPKQMSIPCLVECVNC 1107  
DB 1159 DWSTCOLSEKAVCGNGIKTRMLDCVRSVSDGKVDLYKCEALGLEKNWMTSCMVECPVNC 1218  
QY 1108 QLSGTAWTSCSTCGHGGGMSRTRFIIMPTQGEGRPCPTTELTOEKTCPVTPCYSVLGN 1167  
DB 1219 QLSWSPWSECSQTGTLTGKMIERRITVTPFQDGRPCPSLMDQSKPCPKCYRQYQY 1278  
QY 1168 WSACKLEGCGGQVQIRLSLSCMWHSGSIHAGRVEDALCGEMPF-----QDSILKQLC 1222  
DB 1279 WSPCQVQEAQCCGGRTRNISCVVSDGSADDFSKVDVEEFACADIELIDGNKNMVEESC 1338  
QY 1223 SVPCPGCHLTENSESTCELTICIDGRSFETVGRQSRRTFIIOQSPENODSCPOQVLETR 1282  
DB 1339 SQPCPGDYLKDNSSWSLCLQTCVNGEDLFGGIGVRSRPVILQELENOHLCPQOMLETK 1398

Qy	1283	CTGGGCGCGKGYTEIMKNSGFIDYC-----MKVPGSGEDKKADYGNLSGKNRPNVNSIKHD	1397
Db	1399	SCYDGCQCYEYKKNASAWKSGSSRTTWCORSQDGINVTGGCLVMSQFPDADRSCNPPCPSOPHSY	1458
Qy	1343	CTGGGCGCGKGYTEIMKNSGFIDYC-----MKVPGSGEDKKADYGNLSGKNRPNVNSIKHD	1397
Db	1459	CSEYKTCHEEGTEYVNSSNSTTEQCETLIPVVLFTMEDKRGDKVT-SRAVHPQTPSSNP	1517
Qy	1398	IFKG--WSLOPLDPDGRVKIYWTGVSQGAFLIMIFLIFTSYLVCKKP-KPHQSTPPOQKP	1454
Db	1518	AGRGRWTFLOPPGPDGRLKTYWYGVAAQFVILLIFTVYSMIYLCKKPKKQRRQNTLKP	1577
Qy	1455	LTLAYDGDLDL 1465	
Db	1578	LTLAYDGDADM 1588	

RESULT 8
ABR58450
ID ABR58450 standard; protein; 1598 AA.
XX
AC ABR58450;
XX
DT 07-JUL-2003 (first entry)
XX
DE Human NOV47a.

WPI; 2003-381625/36.  
N-PSDB; ACC72162.

NOVX polypeptides and nucleic acids useful for diagnosing, preventing or treating NOVX-associated disorders, e.g. diabetes, obesity, cancer or dyslipidemia, and in chromosome mapping, tissue typing or pharmacogenomics.

Claim 1; Page 254-255; 487pp; English.

The present invention relates to novel human NOV proteins and their coding sequences (ACC72075-ACC72181 and ABR58363-ABR59469). The NOV proteins are useful in manufacturing a medicament for treating a syndrome associated with a human disease. The NOV proteins and coding sequences may be used to diagnose, treat or prevent metabolic disorders such as diabetes or obesity, infections, cachexia, cancer, neurodegenerative disorders such as Alzheimer's disease or Parkinson's disease, immune disorders, haematopoietic disorders and various dyslipidaemias

Sequence 1588 AA;

Query Match 52.1%; Score 4290; DB 6; Length 1588;  
Best Local Similarity 50.6%; Pred. No. 6.6e-246;  
Matches 765; Conservative 238; Mismatches 450; Indels 58; Gaps 20;

QY 2 VRCIOKLRITVVAKEICEHFALOPPEQACLIIPCRDCCVSEFLPWSNCSGCKGKQLQHR 61  
DB 89 IACIQKQKNDPAEDIIICEYEPKPLEQACLIQCQDCIVSEFSAMWSCSKTCSGGLQHR 148  
QY 62 TRAVIAPPLFGGLQCPNLTESACDAPISCLPGBEEYTFSLKVGPMWSCRILPHLKEINPS 121  
DB 149 TRHVVAPEQFGSGCPNLTEFVCQ---SSPCEAEELRYSLHVGPMWSTCSMPHSRQVRAQ 205  
QY 122 GRT-----VLDPNSSDSNERTVFK-HQSYKAHHHSKSWAIEIGVQTRQVSCTR 167  
DB 206 RRRGNKXERKDRSKGYKQ--PEARLIIKKGNRNQRQENKYWDIIGIQTREVMCMIN 263  
QY 168 SDGQWAMLSLCLQDSFPLTVQSCIMPKDCETQSWSSWSPSCSKTCSGSLLPQFRRSRNNV 227  
DB 264 KTGRAADLSFCQEKLEPMTFQSCVITKEQVSEWSEWSPCKTCHDMVSPAGTVRRTI 323  
QY 228 KMAIIGGKECPLELLEKACIVBELLQOCPRYSWRTSEWKECQVSLLEQDPHHVHTG 287  
DB 324 RQFPTGSEKCEPFEKEPCLSQGDGVPCATYQWRTTEWTECRVDPLLSQDQKRGNGQT 383  
QY 286 PVCGGGQTRREVYCAQS---VFPAALAKAKSVRPVEKALCVGPAFLPSOLCNIPICSTD 343  
DB 384 ALCCGGIQTREVYCVQANENLLSQLSTHKNKASKPMDLKLTGPIPIWTQLCHIPCETE 443  
QY 344 CIVSSWSAWGLCIHENCHEPQGGKGFRTQRHVLMESTGPA---GHCPHLYVESVPCEDPM 400  
DB 444 CEVSPWSAWGPPCTYENCNDQGGKGFKLKRRIITNEPTGGSGVTGNCPLLEAIPCBEPA 503  
QY 401 CYRWLASE-GICPPDHCK-CGLGHRLLKAVCONDRGEDVSGSLC-----PYPPPPERKSC 453  
DB 504 CYDWKAVRLGDCBPDNGKEGPGTQGVVVCINSQGEVDRLCRDAFTPIP-----VAC 558  
QY 454 BIPCRMDCVLSEWTEWSSSCSNKNSDKQTRSTIILALAG-BGGKPCPPSQAQLEHR 512  
DB 559 DAPCEKCVLSTWSTWSSCSHTCSGKTTEGKXIRARSILAYAGEEGGRCFNPSSALQEV 618  
QY 513 LCNHDSQMLHWETSPWGPCSEBTLVTALNATIGWNGEATCGVGTQTRVPCVKSHVQGV 572  
DB 619 SCNEBPTCTVHWQTPGMQGCIEBTVSSFNFTTTWNGEASCSCVGNQTRKVICVRNVQGV 678  
QY 573 MTKKCPDSTPETVRPCFLPCKKDCIVTAFSEWTPCPRMCOAGNATV-KQSRYRIIIIEA 631  
DB 679 GPKKCPSSLAPETVRPCLLPCKKDCIVTYSWTSCTPSCCKEGSSIRKQSRHRVILQLP 738  
QY 632 ANGQOECPTDLTYEERECEDVSLCPVYRWKRPQKWSPCILLPESVWQGITGSSEACGKLQT 691  
DB 739 ANGRGDCITDPIYEEKACEAPQACQSVRWKTHKWRCOLVPSVQODSPQAGEGCGPQROA 798

Y 692 RAVSICSDNRSAMWELCKOTNGMLLVQECTVPCREDCTFTAWSKFTPCSTNCEATKS 751  
b 799 RAITCRKODGGQAGHECLQVAGFPALTAQOIPQDDCQLTSWSPSSCNGDCGAVRT 858  
Y 752 RRRLTJGSRKKEKQDSDLVPLVETELCPDDEFTSQYGNWSDCILPEGRREHRLVR 811  
b 859 RARTLVGSKKKEKCNHLYPLIETQCPCKYNAQPVGNWSDCILPEGVVLLGMKV 918  
Y 812 QADSKECEGLAFRAVACSDKNGRPVDFSCSSGYIOEKVIPCFFDCKLSDWSSWGSC 871  
b 919 QGDIKECGGVYQAMACVDQNGRLVETSRCSHGYYIEACIIPCFSCKLSEWNSRC 978  
Y 872 SSSCGIVRIASKNLKEYPYNGRCPCPLDLX--QVHEAVPCYSECNQYVWVVEHSSC 929  
b 979 SSSCGGVKVRASKWUREKPYNGRCPCPLDLXVNAQVYEVVPCSDCNQYLWVTEPWSIC 1038  
Y 930 KINNELRLSRCCGGTQSRKIRCV-NTADGEGAVDSNLCNDEIPEPTQSCSLMCPNECV 988  
b 1039 KTVFVNRNCGGVYTRKVRQCMQNTADGSEHVEDYLCDEPNPLGSRVCKLPCPEDCV 1098  
Y 989 MSEKGLWCKPSCDPHMQRTTHLR-PSLNSRTCAEDSQVOPCLINENCFOFQNL 1047  
b 1099 ISEWGPWTQCVLPNCQSFQRQADPIRQADDEGRSCENAVEKEPCNLNKCXYHYDNT 1158  
Y 1048 EWSCTQLSENAPCGGVTRILLSCVCSGKPVSDQCEOHNLKEPQMSIPCLVECVNVC 1107  
b 1159 DWSTQLSEKAVCGNGIKTRMLDCVRSQKSDVLYKCEALGLEKNWNTSCWECVPNC 1218  
Y 1108 QLSGWTANTECOTCGHGRMSRTFIIIMPTQGEGRPCPTTELTOEKTCTVPCYSWLGN 1167  
b 1219 QLSWSPWSECSOTCLGCKXIRRTVTQPFQDGRPCPSLMDQSKPCVPCVRYWYQ 1278  
Y 1168 WSAKLEGGDCGEGVQIBSLSCWHSQISHAAGREVEDALCGEMPFF-----QDSILKQLC 1222  
b 1279 WSPQVQAOCEGRTNRNISCVSQSDGADDFSKVDEFCADIELIDGNKNWVLESC 1338  
Y 1223 SVPCGDCHLTEWSEWSTCELTCDIGRSFETVGRQSRRTFIIISFENQDSQPOQVLETR 1282  
b 1339 SOPCPGDCVLDWSSWSLQCLTCVNGEDLFGGQVRSRPVIOLENLQHLCPQOMLETK 1398  
Y 1283 PCTGCKCHYTWKASLWNNERTWQCRSDGVNVTGGSQPARPAAIROCIIPACKRFSY 1342  
b 1399 SCYDGCQYKMAWASAWGSGRTVWCQRSDGINVTGGLVMSQPDADRSCNPPCSQPSY 1458  
Y 1343 CTQGVGCGCEGYEIMKNSGFLDYC-----MKVPGSBDKADVKNI-SGKURPVNSKIH 1397  
b 1459 CSEYTKCHCEGYEVNSNSTLEQCLIPVVLPTMEDKGDVKT-SRAVHTQPSNP 1517  
Y 1398 IFKG--WSLOPLDPGRVKINWYVSGGAFIMIFLFTSYLVCKKP-KPHQSTPPQOKP 1454  
b 1518 AGRGRTWLOPFGDGLKWTWYGAAGAFVLLIPIVSMIYLACKPKPKPQRQNNRLKP 1577  
Y 1455 LTLAYDGLDLM 1465  
b 1578 LTLAYDGDADM 1588  
RESULT 9  
BR58459  
D ABR58459 standard; protein; 1588 AA.  
X C ABR58459;  
X T 07-JUL-2003 (first entry)  
X E Human NOV47j.  
X W Human; NOV; antidiabetic; anorectic; antibacterial; virucide;  
W immunomodulator; cytostatic; nootropic; neuroprotective; dyslipidaemia;  
W antiparkinsonian; antilipaeic; gene therapy; metabolic disorder;  
W diabetes; obesity; infection; cachexia; cancer; Parkinson's disease;  
W neurodegenerative disorder; Alzheimer's disease; immune disorder;

haematopoietic disorder.

Homo sapiens.

W02003029423-A2.

10-APR-2003.

02-OCT-2002; 2002W0-US031358.

02-OCT-2001; 2001US-0326483P.

05-OCT-2001; 2001US-0327342P.

09-OCT-2001; 2001US-0327917P.

09-OCT-2001; 2001US-0328029P.

09-OCT-2001; 2001US-0328044P.

09-OCT-2001; 2001US-0328056P.

12-OCT-2001; 2001US-0328849P.

15-OCT-2001; 2001US-0329414P.

17-OCT-2001; 2001US-0330142P.

22-OCT-2001; 2001US-0341058P.

24-OCT-2001; 2001US-0339266P.

24-OCT-2001; 2001US-0343629P.

29-OCT-2001; 2001US-0349575P.

01-NOV-2001; 2001US-0346357P.

12-APR-2002; 2002US-0371972P.

12-APR-2002; 2002US-0371980P.

17-APR-2002; 2002US-0373261P.

19-APR-2002; 2002US-0373805P.

23-APR-2002; 2002US-0374738P.

16-MAY-2002; 2002US-0381101P.

17-MAY-2002; 2002US-0381635P.

20-MAY-2002; 2002US-0383830P.

01-OCT-2002; 2002US-00262839.

(CURA-) CURAGEN CORP.

Alisbrook JP, Anderson DW, Boldog FL, Burgess CE, Catterton E;

Einger SR, Ellerman K, Gerlach VL, Gorman L, Guo X, Ji W;

Kekuda R, Leach MD, Li L, Miller CE, Patturajan M, Rieger DK;

Rethenbreg ME, Shimkets RA, Smithson G, Spytek KA, Taupier RJ;

Vernet CAM, Voss EZ, Zerhusen BD, Zhong M;

WPI; 2003-381625/36.

N-PSDB; ACC72171.

NOVX polypeptides and nucleic acids useful for diagnosing, preventing or

treating NOVX-associated disorders, e.g. diabetes, obesity, cancer or

dyslipidaemia, and in chromosome mapping, tissue typing or

pharmacogenomics.

Claim 1; Page 264-265; 487pp; English.

The present invention relates to novel human NOV proteins and their coding sequences (ACC72075-ACC72181 and ABR58459). The NOV proteins are useful in manufacturing a medicament for treating a syndrome associated with a human disease. The NOV proteins and coding sequences may be used to diagnose, treat or prevent metabolic disorders such as diabetes or obesity, infections, cachexia, cancer, neurodegenerative disorders such as Alzheimer's disease or Parkinson's disease, immune disorders, haematopoietic disorders and various dyslipidaemias

Sequence 1588 AA;

Query Match 52.0%; Score 4284; DB 6; Length 1588;

Best Local Similarity 50.6%; Pred. No. 1.5e-245;

Matches 764; Conservative 238; Mismatches 451; Indels 58; Gaps 20;

QY 2 VRCIQKLNRTVAVNICEHFALQPTBOACLIPOPCRDVSEFLPWSNCSGCKGLQHR 61

Db 89 IACIQKDXDIPAEIDICEYFEKPLLEQACLIPOCQDCIVSEFSWSECKTCSGLQHR 148

QY 62 TRAVIAPPLFGGLQCPNLNLTESACDAPISCLPGEYEYTFSLKVGWPWSCRLPHLKEINPS 121



149 TRHVAPPQGGGCPNLTFVQVC---SSPCAEELRYSLHVGPNSTCMSPHSRQVRQA 205  
122 GRT-----VLDFNSDSNERVTFK-HOSYKAHHSHKSWAIEIGYOTROVSCTR 167  
206 RRGKKNKEREKDSKVGD---PEARELIKKNRNPQRQENKYWDIQIGYOTREVMCIN 263  
168 SDQONAMLSLCLQDSPLTLVQSCIMPXDCETBSWNSWSPSCSTKTRSGSLLPGRFSRNV 227  
264 KTGKAADLSFCQOEKLPMTFQSVITKECQVSWSEWSPCSKTCHDMVSPAGTRVRTTI 323  
228 KMAIAGGKECPLELLKEACIVGEILLQOCPRYSWRTSEWKECOVSLLEQQDPMHVTG 287  
324 RQPISEKECPFEKEPCLSGQGVVPCATYGNRTTEWTECRVDPLLSQQCKRGNOT 383  
288 PFCGGIQTRETVCAQS---VPAALARAKEYRSRVEKALCVGAPLPSQLCNTPCSTD 343  
384 ALCCGGIQTRETVCVQANENLLSKLTHKNKEASKPMDLKLCTGPITNTTLXCHPCPTE 443  
344 CTVSSSAMGLCIHENCHEPOGKKGRTORHVLVESTGPA---GHCPLVBSVCEBPM 400  
444 CEVSPSAMGCPTEYENCNDXXKKGKKFKRRKRITNEPTGSGVGTGNCPLLEBAICEEPA 503  
401 CYRWLASE-GICFPDGHK-CGLGHRILKAVCQNDGEDVSGSLC-----PVPPPPPERKSC 453  
504 CYDWKAVRLGDCEPDNGKEGPGTQVQEVVVCINSDEEVDRLQCRDAIFPIP-----VAC 558  
454 EICRMDCVLSEWTEWSSCSQSNKNSDKQTRSTILALAG-EGGKPCPPSQALQEH 512  
559 DAPCPKDCVLTSTWSSCSHTCSGKTGKQIRARSILAYAGEEGGIRCPNSSALQEV 618  
513 LCNHSCMQLHWETSPWGPCSEDTLTALNATIGMNGEATCGVIGTRRVFCVKSHVGV 572  
619 SCNEHPCVYVHWQTPNGQCIEDTSVSSNTWTTWNGEASCVSQWQTRKVICRVNVGV 678  
573 MTRCPDSTRTPTVRPCFIPCKKDCIIVTAFSEWTPCPRMQCQAGNAV-KOSRYRIIOEA 631  
679 GPKKCPESLPTVRPCLLPCKKDCIIVTSDWTSFSCPKEDGSSIRKQSHRVIQLP 738  
632 ANGOECPTDLYEERCEDEVLCPVVRWPKQWSPCILVPESWQGITGSSACGKGLQT 691  
739 ANGRDCTDLYIEKACEAPQACQSTRWTKHRRRCOLVPWSVQQSPQAQGGCPGQA 798  
692 RAVSCISDDNRSABMECLKQTMGPLLVQECVTPCREDTFTAMSKFTPCSTNCEATKS 751  
799 RAITCRKQDGGAGIHECLOYAGVPALVQACQIPCODDCQLTSKSKFSSCNGDCGAVT 858  
752 RRBLTKSKRKEKQSDSLYPLVETELCEPDEPI SOPYGNWSDCILPGRPREHRLAV 811  
859 RKRTLVGSKKKKBEKKNSHLYPIETQYCPCKYNAQPVGNWSDCILPBGKVEVLLGMKV 918  
812 QADSKEGGLRFRFVACSDKNGRPVDPFSCSSGYIQEKVIPCFFDCKLSDWSW3SC 871  
919 QGDIKECGQYRYQAMACYDQNGRLVETSRCSNHSYIEEACIIPCSDCKLSEWNSR 978  
872 SSSCGIGVIRSRKWLKXPNGBRCPKLDLKN-QVHEAPVCYSECNOYSWVWEHWS 929  
979 SKSGGKVKVRSKWLREKYPNGRCPKLDHVNQAQVYEVFCHSDCNLYLWTFEWSIC 1038  
930 KINNELRLSCGGTQSRKLRVCY-NTADGEGGAVDSNLNODBIPEPTOSCSLMCNVCV 988  
1039 KVTIVNNRENGEGVQTRKVRQNTADGSEVEDYLCDPEMPGSRVCKLPEDCV 1098  
989 MSEWGLMSKCPQSDPHMTQRRTHLLR-PSLNSRTCAEDSVQPCLLNENCFOFYNL 1047  
1099 ISEGWPTQVCLPCNOSPPQRSADPIRQADREGRCPNAVEKPCNLNKNCHYDYNV 1158  
1048 EWSTCOLSENAPCGGVTRLLSCVSDGKPVSMDOCEQHNLEKPBMSIPCLVRCVNC 1107  
1159 DWSTCQSEKAVONGIKTNLDCVSDGSKVDLKYCBALGLEKNQWMTSCWKECFVNC 1218  
1108 QLSGWTAWTECSQTCGHGRMSRTRFIIMPTQGEGRPCPTELTQETCPWTCYSWLVGN 1167  
1219 QLSWSPSWSECSQTCGLTGKWRRTVTQPFQDGRPCPSLMDQSKPCPVKFCYRWQYQ 1278

QY 1168 WSACKLGGGCGEGVQJRSLSOMVHSGSISHAAGRVEDALCGEMP-----QDSILKQLC 1222  
Db 1279 WSPQVQEAQCGEGTRTNISCVVSDGADDFKVDDEFCADIELIIDGNKMWLEESC 1338  
QY 1223 SVPCGCHLTKSEWSTCBELTICIDGRSFETVGRQSRRTFIIOSEPNQDSCPQVLETR 1282  
Db 1339 SQPCGDCYLKDWSSWSLCLQTCVNGEDLGFGGIQRVSRPVIIQELENOHLCPEQMLETK 1398  
QY 1283 PCTGGKCVHTWTKASLWNNNERTVWCORSQVNVGTCSPQASPAAIRQCIIPACRKPFSY 1342  
Db 1399 SCYDQCYEYKAWASAWKSGSRVWCORSQVNVGTCGLVMSQPDADRSNPPCPSQHSY 1458  
QY 1343 CTQGVGCGCKGKYTEIMKNSGFLDYC-----MKVPGSEDKKADVKNLSGKRRPVNSKIH 1397  
Db 1459 CSETKTCHCEGYTEVMSNSSTLEQCTLIIPVWLPMTMEDKRGDKT-SRAVHPTQPSNP 1517  
QY 1398 IFKQ--NSLOPLDPDGEVKIWWVSGSGLIMFLIFTSYLVCKKP-KPHQSTPPQOKP 1454  
Db 1518 AGRTWFLQPFPGDGLKLTWYVGAAGAFVLLIFIVSMYILACKKPKPKQRQNNRLKP 1577  
QY 1455 LTLAYDGLDM 1465  
Db 1578 LTLAYDGDADM 1588  
RESULT 10  
ABR58453  
ID ABR58453 standard; protein; 1545 AA.  
XX AC ABR58453;  
XX XT  
XX DT 07-JUL-2003 (first entry)  
XX DE Human NOV47d.  
XX KW Human; NOV; antidiabetic; anorectic; antibacterial; virucide;  
KW immunomodulator; cytostatic; nootropic; neuroprotective; dyslipidaemia;  
KW antiparkinsonian; antilipemic; gene therapy; metabolic disorder;  
KW diabetes; obesity; infection; cachexia; cancer; Parkinson's disease;  
KW neurodegenerative disorder; Alzheimer's disease; immune disorder;  
KW haematopoietic disorder.  
XX OS Homo sapiens.  
XX PN WO2003029423-A2.  
XX PD 10-APR-2003.  
XX PF 02-OCT-2002; 2002WO-US031358.  
XX PR 02-OCT-2001; 2001US-0326483P.  
PR 05-OCT-2001; 2001US-0327342P.  
PR 09-OCT-2001; 2001US-0327917P.  
PR 09-OCT-2001; 2001US-0328029P.  
PR 09-OCT-2001; 2001US-0328044P.  
PR 09-OCT-2001; 2001US-0328056P.  
PR 12-OCT-2001; 2001US-0328849P.  
PR 15-OCT-2001; 2001US-0329414P.  
PR 17-OCT-2001; 2001US-0330142P.  
PR 22-OCT-2001; 2001US-0341058P.  
PR 24-OCT-2001; 2001US-0339266P.  
PR 24-OCT-2001; 2001US-0343629P.  
PR 29-OCT-2001; 2001US-0349575P.  
PR 01-NOV-2001; 2001US-0346357P.  
PR 12-APR-2002; 2002US-0371972P.  
PR 12-APR-2002; 2002US-0371980P.  
PR 17-APR-2002; 2002US-0373261P.  
PR 19-APR-2002; 2002US-0373805P.  
PR 23-APR-2002; 2002US-0381101P.  
PR 16-MAY-2002; 2002US-0381635P.  
PR 17-MAY-2002; 2002US-0381635P.  
PR 29-MAY-2002; 2002US-0383830P.

PR	01-OCT-2002; 2002US-00262839.	Db	619	SCNEHPCVTHVQTPGWPWGQCIEDTSVSSFNTHWNGEASCSVGMQTRKVICVRVNVQV	678
XX	(CURA-) CURAGEN CORP.	Qy	573	MTKECPDSTRPETYRCPFLPCKKDCIYTABSEWTPCPRMQAGNATV-KOSRYEIIQEA	631
PA	Alisbrook JP, Anderson DW, Boldog FL, Burgess CE, Catterton E;	Db	679	GPKCPDSTRPETYRCPFLPCKKDCIYTABSEWTPCPRMQAGNATV-KOSRYEIIQEA	738
PI	Edinger SR, Ellerman K, Gerlach V, Gorman L, Guo X, Ji W;	Qy	632	ANGGQECFDTLYEERECEDVSLCFVYRKQKQKSPFCILVPESVWQGITGSEACGKLQT	691
PI	Rekuda R, Leach MD, Li L, Miller CE, Patturajan M, Rieger DJ;	Db	739	ANGGRDCTDPLYEEKACEAPQACQSYRKWKHKRRCCQLVPSVQDQSPGAGCGGGRQA	798
PI	Rothenberg ME, Shimkets RA, Smithson G, Spytek KA, Taupier RJ;	Qy	692	RAVSCISDDNRSAEMMECLKQTNMGMLLVQECVPCREDCTFFAWSKETPCSTCEATKS	751
PI	Vernet CAM, Voss EZ, Zerhusen BD, Zhong M;	Db	799	RAITCRKQDQGGQAGIHECLQYAGVPALQACQPCQDDCQLTWSKFSFSCNGCGAVRT	858
XX	WPI: 2003-381625/36.	Qy	752	RRRLQTKSRKKEKQSDLDYPLVETELCPDEFISQPYGNWSDCILPEGRREPHRLVR	811
DR	N-PSDB; ACC72165.	Db	859	RKRTLVGSKKKKCKNSHLYPLIETQYCPCKYNAQPVGNWSDCILPEGKVEVLGMKV	918
PT	NOVX polypeptides and nucleic acids useful for diagnosing, preventing or	Qy	812	QADSKCEGELRFRFAVACSKDKNGRPVDPSPCSSSGYIOEKVIPCPCPDCKLSDSSWGSC	871
PT	treating NOVX-associated disorders, e.g. diabetes, obesity, cancer or	Db	919	QGDICEGQGYRYQAMACYDQNGRLVETSRCSHGYIEEACIIPCPSDCKLSEWNSMRC	978
PT	dyslipidemia, and in chromosome mapping, tissue typing or	Qy	872	SSSGIGVIRIRSKWLKPKYNGRPPCPKLDLN--QVHEAVPCYSECNOXSVVVEHSSC	929
PT	pharmacogenomics.	Db	979	SKSCGSGVYKRSKWLKPKYNGRPPCPKLDHVNQAQYVYVPCSDCNQYLWTEPHSIC	1038
XX	Claim 1; Page 257; 487pp; English.	Qy	930	KINNELSLRCGGTQSRKIRCV-NTADGEGAVDSMLCNQDEIIPPTQSCSLMCPNECV	988
XX	The present invention relates to novel human NOV proteins and their	Db	1039	KVTFVNRKENGEGVQTRKVRQMONTADGSEHVEDYLCDEEPLGSRVCKLPCPDECV	1098
CC	coding sequences (ACC72075-ACC72181 and ABR58469). The NOV	Qy	989	MSEGLMSKCPQSCDPTMQPRTREHLR-PSLNSRTCAEDSOVOPCLLNENCFQYNLT	1047
CC	proteins are useful in manufacturing a medicament for treating a syndrome	Db	1099	ISEWGPWTQCVLPNCQSFRRORSADPIRQFADGSRSCPNAVEKEPCNLNKNCHYDNTV	1158
CC	associated with a human disease. The NOV proteins and coding sequences	Qy	1048	EWSQCQLSENAAPCGQVYRTRLLSCVCSGDKPVSMDQCEQHNLEKPPQMSIPCLVECVVNC	1107
CC	may be used to diagnose, treat or prevent metabolic disorders such as	Db	1159	DWSTCQLSEKAVCGNGIKTRMLDVRSDGKSDLYKCEALGLEKNWQNTSCWVECPVNC	1218
CC	diabetes or obesity, infections, cachexia, cancer, neurodegenerative	Qy	1108	QLSQTAWTECSQTCGHHGRMSRTRFIMPTQEGRGRCPTLTQKTCPTVTPCYSWLGN	1167
CC	disorders such as Alzheimer's disease or Parkinson's disease, immune	Db	1219	QLSDWSPMSECSQTCGLTKMIRERTVTPFQDGRPCPSLMDQSKPCPKVPCYRWYQG	1278
CC	disorders, haematopoietic disorders and various dyslipidaemias	Qy	1168	WSACKLEGDCGEGVQIRSLSCMVHSGSISHAAGRVEDALCDEMPF---QDSILKQLC	1222
XX	Sequence 1545 AA;	Db	1279	WSPQVQEAQCGEGTRTENTSCVSDGSADDFSKVVDDEFCADIELIIDGNKNVLESC	1338
Query Match	49.98; Score 4113.5; DB 6; Length 1545;	Qy	1223	SVPCGDCHLTWSGEMSTCELTIDGRSFETVGRQSRRTFIQSFEQNSCPOQVLETR	1282
Best Local Similarity	50.34; Pred. No. 1.9e-235;	Db	1339	SQPCPGDCYLKDWSSWSLQCLTCVNGEDLGFGGIQRVSRPVIQELNHLCPQMLETK	1398
Matches 732; Conservative 229; Mismatches 436; Indels 57; Gaps 19;		Qy	1283	PTGCKCVHYTWKASLWNNERTVWCORSQSDGVNVTGCSQPARAAIROCIIPACKRPPSY	1342
2 VRCIQKLNRTWANEICEHPALOPPEEQACILPCPRDCVVSSEFLPMSNCKGCKKLOHR	61	Db	1399	SCYQGCQVEYKMAKASAKWGSRTVWCORSQSDGVNVTGCLVMSQDADRSCNPPCSQPHSY	1458
89 IACTQKDKDPAEDIIICEYEPKPLEEQACILPCQDCIVSEFSAMSECKTCGSLQHR	148	Qy	1343	CTQGGVGCCEKGYTEIMKSNGLDYC-----MKVPGSGEDKADVNLNKGKRRPVNSKIH	1397
62 TRAVIAPPLPGGLCCNLTSTRACDAPISCPLEEBEYTPSLKVPMSKRLPHLKEINPS	121	Db	1459	CSETKTCHEEGYEVWSSNSTLEOCTLIPIVWLPMTMEDKRGDVKT-SRAVHPQTQSSNP	1517
149 TRHVAVAPPQGGGCPNLTEFQVC---SFCABEELRYSLHVGPMSTCSMPESRQVQA	205	Qy	1398	IFKG--WSLQPLDP 1409	
122 GRT-----VLDNPSNSRNVTPK-HOSYKAHHSSKSWAIEIGYOTROVSCTR	167	Db	1518	AGRGTWFLQPFPG 1531	
206 RRGKNKEREKRSKGVKD--PEARLIIKKRNKNQKNQENKYWDIQLGYQTRVMCIN	263	Qy	RESULT 11		
168 SDGQAMLSLCLQDSPLTVQSCIMPCKDCQTSQWSSWSPCKTCSRGSLLPGFRSRNV	227	Db	ABR58455		
264 KTGKAADLSFCQCKLPMTPQSCVITKECQVSEWSEWSPCKTCHDMVSPAGTRVTRTI	323	Qy	ID ABR58455		
228 KMAIGGKCEPCLLEKEACIVGELLQOCPRYSWETSEWKECVSLLEQODPHWHTG	287	Db	AC ABR58455;		
324 RQPIGEKCEPFEKPECLSGDGVVPCATVGMWTTTEWTECRVPLLSQQDKRGNOT	383	Qy	DT 07-JUL-2003 (first entry)		
288 PVCGGGIQTREVTCAOS-----VPAALRAKESRVERKALCVGAPLPSQLCNIPCSID	343	Db	DE Human NOV47f.		
384 ALCGGGIQTREVTVCQANENLLSQLSHKKNKASKPMDLKLCITGPIPTNLQCHIPCPE	443	Qy	XX		
344 CIVSSWSAWGLCHENHEPQGGKGRTRORHVLMESTGPA---GHCPLHVSVPCEDEPM	400	Db	XX		
444 CEVSPWSAWGPTCYENCNDQQKGGFKLRRTTNEPTGGSGVTGNCPLHLEAIPCEEPA	503	Qy	XX		
401 CYRWLASE-GICFPDGHK-CGLGHRILKAVCONDRGEDVSGSLC-----PVPPPPKXSC	453	Db	XX		
504 CYDWKAVRLGDCFPDNGKEGPGTQVQVVCINSDEGVEDRQLCRDAIFIP-----VAC	558	Qy	XX		
454 EIPCRMDCVISEWTEWSSCSQSNKNSDKQTRRTILALAG-EGKPCPPSQALQEH	512	Db	XX		
559 DAPCPKDCVLTSTWSTWSSCSHTCSGKTEGQIRARSILAYAGEGGIRCPNSSLQEV	618	Qy	XX		
513 LCNDHSCWQJLHWETS PMGPCSEDTLTALNATIGWNGEATCGVIGIOTRIVFCVSHVGOV	572	Db	XX		

KW Human; NOV; antidiabetic; anorectic; antibacterial; virucide;  
 KW immunomodulator; cytotatic; nootropic; neuroprotective; dyslipidaemia;  
 KW antiparkinsonian; antilipemic; gene therapy; metabolic disorder;  
 KW diabetes; obesity; infection; cachexia; cancer; Parkinson's disease;  
 KW neurodegenerative disorder; Alzheimer's disease; immune disorder;  
 KW haematopoietic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003029423-A2.  
 XX  
 PD 10-APR-2003.  
 XX  
 PF 02-OCT-2002; 2002WO-US031358.  
 XX  
 PF 02-OCT-2001; 2001US-0326483P.  
 XX  
 PR 05-OCT-2001; 2001US-0327342P.  
 XX  
 PR 09-OCT-2001; 2001US-0327917P.  
 XX  
 PR 09-OCT-2001; 2001US-0328029P.  
 XX  
 PR 09-OCT-2001; 2001US-0328044P.  
 XX  
 PR 09-OCT-2001; 2001US-0328058P.  
 XX  
 PR 12-OCT-2001; 2001US-0328849P.  
 XX  
 PR 15-OCT-2001; 2001US-0329414P.  
 XX  
 PR 17-OCT-2001; 2001US-0330142P.  
 XX  
 PR 22-OCT-2001; 2001US-0341058P.  
 XX  
 PR 24-OCT-2001; 2001US-0339266P.  
 XX  
 PR 24-OCT-2001; 2001US-0343629P.  
 XX  
 PR 29-OCT-2001; 2001US-0343575P.  
 XX  
 PR 01-NOV-2001; 2001US-0348357P.  
 XX  
 PR 12-APR-2002; 2002US-0371972P.  
 XX  
 PR 12-APR-2002; 2002US-0371980P.  
 XX  
 PR 17-APR-2002; 2002US-0373261P.  
 XX  
 PR 19-APR-2002; 2002US-0373805P.  
 XX  
 PR 23-APR-2002; 2002US-0374738P.  
 XX  
 PR 16-MAY-2002; 2002US-0381101P.  
 XX  
 PR 17-MAY-2002; 2002US-0381635P.  
 XX  
 PR 29-MAY-2002; 2002US-0383830P.  
 XX  
 PR 01-OCT-2002; 2002US-00262839.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 PI Alsobrook JP, Anderson DW, Boldog FL, Burgess CE, Catterton E;  
 PI Edinger SR, Ellerman K, Gerlach VL, Gorman L, Guo X, Ji W;  
 PI Kekuda R, Leach MD, Li L, Miller CE, Patturajan M, Rieger DK;  
 PI Rothenberg ME, Shinkets RA, Smithson G, Spytek KA, Taupier RJ;  
 PI Vernet CAM, Voss EZ, Zerhusen BD, Zhong M;  
 XX  
 DR WPI: 2003-381625/36.  
 DR N-PSDB; ACC72167.  
 XX  
 PT NOVX polypeptides and nucleic acids useful for diagnosing, preventing or  
 PT treating NOVX-associated disorders, e.g. diabetes, obesity, cancer or  
 PT dyslipidemia, and in chromosome mapping, tissue typing or  
 PT pharmacogenomics.  
 XX  
 PS Claim 1; Page 260; 487pp; English.  
 XX  
 CC The present invention relates to novel human NOV proteins and their  
 CC coding sequences (ACC72075-ACC72181 and ABR58363-ABR58465). The NOV  
 CC proteins are useful in manufacturing a medicament for treating a syndrome  
 CC associated with a human disease. The NOV proteins and coding sequences  
 CC may be used to diagnose, treat or prevent metabolic disorders such as  
 CC diabetes or obesity, infections, cachexia, cancer, neurodegenerative  
 CC disorders such as Alzheimer's disease or Parkinson's disease, immune  
 CC disorders, haematopoietic disorders and various dyslipidaemias  
 XX  
 SQ Sequence 1549 AA;  
 XX  
 Query Match 49.7%; Score 4095.5; DB 6; Length 1549;  
 Best Local Similarity 50.2%; Pred. No. 2.2e-234;  
 Matches 730; Conservative 230; Mismatches 437; Indels 57; Gaps 19;  
 XX 2 VRCIOKLNRTVVANBICEHFALQPPTQACLIQPCRDVCVSEFLFWSNCSKCGKGLQHR 61

Db 91 IACIQKDXDIPAEIICIEYFEKPLLEOACLIQCOQDQIVSEFSWSECKTSGSLQHR 150  
 Qy 62 TRAVIAPPLFGLQCPNLTESACDAPISCPUGSEYFTSLKVGWPKRLPHLKEINPS 121  
 Db 151 TRHVAPPOFGSGGCPNLTQFVCO---SSPCEAEELRYSLHVGWPKSCMPSHQVRQA 207  
 Qy 122 GRT-----VLDNDSNERTVFK-HQSYKAHHHSKSWAIEGYQTRQVSCTR 167  
 Db 208 RRRGKNKEREKDRSGVKD--PEARELIKKNRNRQRNQENKYWDIQIQTREVMCIN 265  
 Qy 168 SDGONAMLSLICLDSPFLTVQSCIMPKOCETQSMWSWSPCSKTSGSILLPGFRSRNV 227  
 Db 266 KTGKAADLSFCQOEKLPMTFQSCVITKEQVSENERSPCSKTCHDMVSPATRVRTI 325  
 Qy 228 KXMAIGGKECPLELEKEACIVEGELLQOCPRYSWRTSEWKEQVSLLEQQDPHHVYG 287  
 Db 326 RQPPIGSEKEPEFEKEPCLSQDGVWPCATYGWRTTEWTECRVDPDLSQDKEGNGT 385  
 Qy 288 PYCGGGIOTREYVCAQS-----VPAALALRAKEVSPPEKALCVGPAPLPOLCNTPCSTD 343  
 Db 386 ALCGGGIOTREYVCOANENLLSQJSTHKNKEASKPMDLKLCCTGPIPNMTQJCHIPCPT 445  
 Qy 344 CIVSSWSAMGLCIHENCHEPOGKGFRTQRQHVLMESTGPA---GHCPLHVESVCEPDM 400  
 Db 446 CEVSPWSANGPCTYENCNDQOGKGFKLKRRIITNEPTGGSGVTGNCPHLEAIPCEBPA 505  
 Qy 401 CYRWLASE-GICFPDHGK-CGLGHRILKAVCONDRGEDVSGSLC-----PVPPPERKSC 453  
 Db 506 CYDMKAVRLGNCEPDNGKEGPGTQVQEVVVCINSQGEVDRQLCRDAIFPI- - - - -VAC 560  
 Qy 454 EIPCRMDCVLSWETWSSQSCSNKNSDKGTRTIRTIILALAG-EGGKPCPPSQALQHR 512  
 Db 561 DAPCPKCVLSTWSTWSSCSHTCSGKTIEGKQIRARSILAYAGEGGIRCPNSSALQEV 620  
 Qy 513 LNDHSCMQLHWETSPWGPCSEDTLVTALNATIGNGEBATCGVGTQTRVFCVKSHVGV 572  
 Db 621 SCNEHPTVYHMQTGPWGQCIEDTYSSEFNITTTWNGEASCSVGMQTRKVICRVNVGV 680  
 Qy 573 MTKRCPDSTRPETVRPCPLPCKKCKQINGMPLLVQECTVPCREDCTFTAMSKETPCSTNCEATKS 751  
 Db 681 GPKCPESLRPETVRPCLLPCKKECIVTPYSWTSCTPSCKEGDSIRKQSRHVIQLP 740  
 Qy 632 ANGQCEPDTLYEERECEPDVSLCPVYRWKPKQWSPCIIIPVPSWQGITGSSSEACGKQT 691  
 Db 741 ANGGRDCTPLYEERACEAPQACQSVYRWKTHKRRKQCLVPWSVQDQSPGAQEGCGPGRQA 800  
 Qy 692 RAVSCISDDNRSANEMMELCKQINGMPLLVQECTVPCREDCTFTAMSKETPCSTNCEATKS 751  
 Db 801 RAITCKRQGGQAGHIECLQYAGVPALTAQCIQDQDQJTSWKFSSCNGDCGAVRT 860  
 Qy 752 RRRQLTGSRKEKQCDSDLYPLVETELCPDCEFTISQYGNWSDCILPEGREPHRLVR 811  
 Db 861 RKEITLVGSKSKKCKKNSHLYPLIETQVCPDKYNAQPVGNWSDCILPEGVVLLGMKV 920  
 Qy 812 QADSKECEGLRPRAVACSDKNGRPVDPSPFCSSSYIQEKVIPCPFPCKLSDNWSWSC 871  
 Db 921 QGDIKECGGYRYQAMACYDQNGRLVETSRCSHGYIEACIIPCPDCKLSEWNSRRC 980  
 Qy 872 SSSCGIGVNRIRSKWLKEKPYNGRCPCKLDLKN--QVHEAVPCYSECNQYSWVWVHWSSC 929  
 Db 981 SKSCGSGVVRKWLREKPYNGRCPCKLDHVNQAVYEVVPCVHSDCNQYLWVTPWISIC 1040  
 Qy 930 KINNELSRURCGGTQSRKIRCV-NTADGEGAVDSNLQNDQEIIPETQSCSLMCPNECV 988  
 Db 1041 KVTVMNRENCGEVQTRKVRQWNTADGSEHVEDYLCDPPEMPLGSRCKLPCPEDCV 1100  
 Qy 989 MSEWGLWSKSCPOSQDPTMQRTRHLR-PSLNSRTCAEDSOVQCLLNENCFQYQNL 1047  
 Db 1101 ISEWGPWTCVTPCQOSSPQRSADPIRQPADEGRSCFNVAKEFCNLKNKYHYDNYVT 1160  
 Qy 1048 EWSCTQSLSENAPCGGVTRTLSSVCSDGKPVSMQDQCEQHNLEKFORMSIPCLVECVVNC 1107

```

1161 DWSTCOLSEKAVCGNGIKTRMLDCVRSKDGKVDLKYCEALGLEKNQWNTSCWVECPVNC 1220
1108 QLSGWTATWTECSTCHGGRMSTRRLIMPOTGEGRPCEPTELTQKTCFVTPCYSWVLGN 1167
1221 QLSDFSWSECSCTGLTGKMIARRTVTOPFGDGRPCPSLMDQSKPCFVPCYRWQYGO 1280
1168 WSACKLEGGDCGEGVQIRSLSCWVHSGSISHAAGRVEDALCGEMPF-----QDSILKQLC 1222
1281 WSPCQVQEAQCGEGTRNRISCVSDGSDADFSKVDFEFCADIELIIDGNKNWVLESC 1340
1223 SVPCPDCHLTWSEWSTCELTICDGRSPETVGRSRSTFIIOSEFQDSCPOQVLSTR 1282
1341 SPCPCGDCVLKDWSSWLSCLTGVAGEDLGFGGIVQVRSRPVILQELNQHLCPEQMLETK 1400
1283 PCTGGKCYHYTWKASLWNNNERTVWCQRSDGVNVTGGCSPOARPAARQCIACPKRPTSY 1342
1401 SCVDGQCYEYKMWASAKGSSRTVWCQRSDGINVTGGCLVMSQPADRSCHNPPCSQPHSY 1460
1343 CTQGGVCGCEKGYTEIMKNSGFLDYC-----MKVPGSEBKKADVKNLKGKORPVNSKIHD 1397
1461 CSETKTCHCEEGYTEVMSNSTLEOCTLIPVVVVLPTMEDKRGDVKT-SRAVHPTQPSNP 1519
1398 IFKG--WSLOPLDP 1409
1520 AGRGRTWFLQPGP 1533

RESULT 12
ABR58454
[CD ABR58454 standard; protein; 1490 AA.
AC ABR58454;
AT 07-JUL-2003 (first entry)
AE Human NOV47e.
X Human; NOV; antidiabetic; anorectic; antibacterial; virucide;
W immunomodulator; cytostatic; nootropic; neuroprotective; dyslipidaemia;
W antiparkinsonian; antipapemic; gene therapy; metabolic disorder;
W diabetes; obesity; infection; cachexia; cancer; Parkinson's disease;
W neurodegenerative disorder; Alzheimer's disease; immune disorder;
W haematopoietic disorder.
AS Homo sapiens.
X WO2003029423-A2.
X 10-APR-2003.
X 02-OCT-2002; 2002WO-US031358.
X 02-OCT-2001; 2001US-0326483P.
X 05-OCT-2001; 2001US-0327342P.
X 09-OCT-2001; 2001US-0327917P.
X 09-OCT-2001; 2001US-0328029P.
X 09-OCT-2001; 2001US-0328044P.
X 09-OCT-2001; 2001US-0328056P.
X 12-OCT-2001; 2001US-0328849P.
X 15-OCT-2001; 2001US-0329414P.
X 17-OCT-2001; 2001US-0330142P.
X 22-OCT-2001; 2001US-0341058P.
X 24-OCT-2001; 2001US-0343266P.
X 24-OCT-2001; 2001US-0343629P.
X 29-OCT-2001; 2001US-0343575P.
X 01-NOV-2001; 2001US-0345357P.
X 12-APR-2002; 2002US-0371980P.
X 12-APR-2002; 2002US-0371972P.
X 17-APR-2002; 2002US-0371980P.
X 19-APR-2002; 2002US-0373261P.
X 23-APR-2002; 2002US-0373805P.
X 16-MAY-2002; 2002US-0374738P.
X 17-MAY-2002; 2002US-0381101P.
X 17-MAY-2002; 2002US-0381635P.

```

```

29-MAY-2002; 2002US-0383830P.
01-OCT-2002; 2002US-00262839.
(CURA-) CURAGEN CORP.
Alsobrook JP, Anderson DW, Boidog FL, Burgess CE, Catterton E;
Edinger SR, Ellerman K, Gerlach VL, Gorman L, Guo X, Ji W;
Kekuda R, Leach MD, Li L, Miller CE, Patturajan M, Rieger DK;
Rothenberg WE, Shinkets RA, Smithson G, Spytek KA, Taupier RO;
Vernet CAM, Voes EZ, Zerhusen BD, Zhong M;
WPI; 2003-381625/36.
DR N-PSDE; ACC72166.
XX
NOVX polypeptides and nucleic acids useful for diagnosing, preventing or
treating NOVX-associated disorders, e.g. diabetes, obesity, cancer or
dyslipidaemia, and in chromosome mapping, tissue typing or
pharmacogenomics.
Claim 1; Page 258-259; 487pp; English.
XX
The present invention relates to novel human NOV proteins and their
coding sequences (ACC72075-ACC72181 and ABR58363-ABR58469). The NOV
proteins are useful in manufacturing a medicament for treating a syndrome
associated with a human disease. The NOV proteins and coding sequences
may be used to diagnose, treat or prevent metabolic disorders such as
diabetes or obesity, infections, cachexia, cancer, neurodegenerative
disorders such as Alzheimer's disease or Parkinson's disease, immune
disorders, haematopoietic disorders and various dyslipidaemias
XX
Sequence 1490 AA;
SQ
Query Match 48.9%; Score 4033; DB 6; Length 1490;
Best Local Similarity 48.1%; Pred. No. 1.1e-230;
Matches 724; Conservative 224; Mismatches 414; Indels 142; Gaps 18;
Qy 2 VRGLOKLRVTVANEICEHPALOPPEOACLPDPCDVSEFLPWSNCSKGGKGLQHR 61
Db 89 IACIQKDQKIDPAEDIIICEYFEPPELLEOACLPDPCDVSEFLPWSNCSKGGKGLQHR 148
Qy 62 TRAVIAPPLFGGLQCPNLTESRACDAPISCPICEEETYSKVGPMKCRLLPKLKNPS 121
Db 149 TRHWVAPPQGGSGCPNLTEFQVCQ---SSPCEAEELRYSLHVGPMSTCSMPHSRQVQA 205
Qy 122 GRT-----VLDNSDSNERTVFK-HQSYKAHHSHKSWAIEIGYQTRQVSCTR 167
Db 206 RRRGKXKEREKDRSKGVKD--PEARELIKKNRNRQENKYNWDIQGYQTRVYMCIN 263
Qy 168 SDGONAMLSLCLQDSPLTVQSCIMPKDCETSWSSWSPSCSKTCRSGSLLPGRSSRN 227
Db 264 KTGKAADLSFCQEQELPMTFQSCVITKEQVSESWSPSKTCHDMVSPAGTRVETRI 323
Qy 228 KHAIGGGKECPLELKEACIVEGELLQCPRYSWRTSEWKECQVSLLEQDPPHVVGT 287
Db 324 RQPPIGSEKECEPEFEKEPCLSQDGVWPCATYGTWRTTEWTCRVDPLLSQDQKRGNT 383
Qy 288 FVCGGGIQTREVVYCAOS----VPAALRAKEVSRVEKALCVGPAPLSQLCNIPCST 343
Db 384 ALCGGGIQTREVVYCAQANENLLSQLSTHKNKEASKPMDLKLCTGPIPTQLCHIPCPE 443
Qy 344 CIVSSWSAWGLCIHENCHEPQGGKGRTRQRHYLMESTGPA---GHCPLHVESVPCEDPM 400
Db 444 CEVSPWSAWGFCITYENCNDQOGKKGFKLRKRITNEPTGGSGVGTGNCPLHLEAIPC 503
Qy 401 CYRWLASE-GICFPDHGK-CGLGHRILKAVQNDRGEDVSGSLC-----PVPPPERKSC 453
Db 504 CYDWKAVRLGDCPEPDNGKEGCGPTQVEVVCINSDEGEVDRLCDRAIFPIP-----VAC 558
Qy 454 EIPCRMDCVLSEWTSWSSQSCSNKNKSGKOTRSRTILALAG-EGGKPCPPSOALQEH 512
Db 559 DAPCPKDCVLSWTSWSSSHCHSCGKTGKQIRASILAYAGEEGGICPNSSALXEV 618
Qy 513 LCNDHSCMQLHWETSPWGPCSEDTLVTALNATIGWNGEATCGVGIGTRVFCVKSHGVQ 572

```

```
db 119 SCNEHPCVTHXGTGPGXCIETDTSVSSFTTTTNGEASCSVGMXTRKVICVRNVGV 678
2y 573 MTRKCPDSTPRTVPCFIPCKKDCIATVAFSEWTPCFRMCQAGNAV-KOSRYRIIIEA 631
db 679 GPKKCPESLAFETVRPCLLIPCKKDCIATVAFSEWTPCFRMCQAGNAV-KOSRYRIIIEA 738
2y 632 ANGQECFDTLYEERECEDVSLCPVVRWPKQWSPGILVPESVWQGITSSACGKGLQT 691
db 739 ANGRCCTDLYEERECEDVSLCPVVRWPKQWSPGILVPESVWQGITSSACGKGLQT 798
2y 692 RAVSCISDDNRSABMECLKQTMGELLVQECTVPCREDCTFTANSKFTPCSTNCEATKS 751
db 799 RAITCRKQDGGQAGIEHCLQYAGVFPALTOACGIPQDDCQLTWSKFSKSCNGDCGAVRT 858
2y 752 RRRLGTGSKSEKCKODSDLYPLVETELCPDEFISQPYGNWSDCLTPEGREPHGLRV 811
db 859 RKTTLVGSKKKCKKNSHLYPIETQYCPDKYNAQPVGNWSDCLTPEGREPHGLRV 918
2y 812 QADSKEGGLRFRVACSDKNGRPVDPFSCSSGYIOEKCVIPCFDCKLSDWSSWGC 871
db 919 QGDIKECGQYRYQACVYDQNGRLVETSRCSHGVEEACIIPCFSDCKLSEWSNWC 978
2y 872 SSSCGIVRTRSKWLKPKNGRCPCKLDLKN-QVHEAVPCYSECNQYSWVEHWSSC 929
db 979 SKCGSGVKVRSKWLREKPYNGRCPCKLDLKN-QVHEAVPCYSECNQYSWVEHWSSC 1038
2y 930 KINNELSLRCGGTQSRKIRCV-NTADGGGAVDNLQNDIPIPETQSCSLMPCNECV 988
db 1039 KVTFVNNRENGEGVQTRKVRCKMONTADGSEHVEDYLCDEMPVLSRVCKLPCFEDCV 1098
2y 989 MSEWGLWSKCPQSCDPTMORRTHLLR-PSLNSRTCAEDSQVOPCLNENCFOYNLT 1047
db 1099 ISEKPTQVCLPCNQSSFPKQADPIQADREGRCPNAVEKPCNLKNCYHYDNYVT 1158
2y 1048 EWSSTCOLSENAPCGQGVTRLLSCVSDGKPVSMDOCEQHNLEKPKQMSIPCLVECVNC 1107
db 1159 DASTCOLSEKAVCGNGIKTRMLDCVRSKSDGKPVSMDOCEQHNLEKPKQMSIPCLVECVNC 1218
2y 1108 QLSGWTATWTCSTCGHGRMSRTRFIIMPTQGEGRPCPTLTQETKTCPTVPCYVWLN 1167
db 1219 QLSGWTATWTCSTCGHGRMSRTRFIIMPTQGEGRPCPTLTQETKTCPTVPCYVWLN 1278
2y 1168 WSACKLGGGCGEGVQIRSLSCMVHSGSI SHAAGRVEDALCGEMPF-----QDSILKOLC 1222
db 1279 WSPQVQEAQCGEGRTRNRSICVVDGSDGADDFKVDVEEFCADELIIDGNKNWLESC 1338
2y 1223 SVPCPGCHLTENSEWSTCELTICIDGRSFETVGRQSRRTFIIQSFENQDSCPQVLETR 1282
db 1339 SOPCPGDCYLKDWSSWSLQCLTCVNGEDLFGGIGVRSRPTVIOELNHLCPQMLETK 1398
2y 1283 PCTGGKCVHYTWKASLWNNERTVWCORSQGVNVTGCSQPARPAATROCIACRKPFSY 1342
db 1399 SCYQGCYKWKASAWKSGSRVWCORSQGVNVTGCSQPARPAATROCIACRKPFSY 1433
2y 1343 CTQGGVCGCEKGYTEIMKSNGLFDYCMKVPGESEDKADVNKLSGKNRPVNSKIHDFKGV 1402
db 1434 -----DGRUKTWYGYAAGAFVLLIFTVSMYIACKPKPKPQRQNNRLKPLTLAYDG 1486
2y 1462 DLDM 1465
db 1487 DADM 1490
```

RESULT 13  
AAO15279  
ID AAO15279 standard; protein; 484 AA.  
XX  
AC AAO15279;

```
XX 16-SEP-2002 (first entry)
DT Human large protein 53-24.
XX Human; large protein 53.24; embryonic development deformity;
XX protein metabolism disturbance; tumour; immunologic system disturbance.
XX Homo sapiens.
XX CN1333246-A.
XX 30-JAN-2002.
XX 07-JUL-2000; 2000CN-00117029.
XX 07-JUL-2000; 2000CN-00117029.
XX (SHAN-) SHANGHAI BIODOOR GENE DEV CO LTD.
XX Mao Y, Xie Y;
XX WPI; 2002-305572/35.
XX N-PSDB; AAL43703.
XX New human large protein 53.24 and encoding polynucleotide, useful for
XX treating embryonic development deformity, protein metabolism disturbance,
XX tumor and immunologic system disturbance disease.
XX Claim 1; Page 29-30 (Disclosure); 36pp; Chinese.
XX The invention comprises the amino acid and coding sequence of the human
XX large protein 53.24. The 53.24 DNA and protein sequences of the invention
XX are useful for treating embryonic development deformity, protein
XX metabolism disturbance, tumour and immunologic system disturbance. The
XX present amino acid sequence represents the human large protein 53.24
XX
XX Query Match 33.1%; Score 2729; DB 5; Length 484;
XX Best Local Similarity 99.8%; Pred. No. 7.7e-154;
XX Matches 483; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 982 MCPNECVNSELGWSKCPQSCDPTMORRTHLLRPSLNSRTCAEDSQVOPCLNENCFO 1041
DB 1 MCPNECVNSELGWSKCPQSCDPTMORRTHLLRPSLNSRTCAEDSQVOPCLNENCFO 60
QY 1042 FOYNLTENSTCOLSENAPCGQGVTRLLSCVSDGKPVSMDOCEQHNLEKPKQMSIPCLV 1101
DB 61 FOYNLTENSTCOLSENAPCGQGVTRLLSCVSDGKPVSMDOCEQHNLEKPKQMSIPCLV 120
QY 1102 ECVNVCQLSGWTATWTCSTCGHGRMSRTRFIIMPTQGEGRPCPTLTQETKTCPTVPCY 1161
DB 121 ECVNVCQLSGWTATWTCSTCGHGRMSRTRFIIMPTQGEGRPCPTLTQETKTCPTVPCY 180
QY 1162 SVNLGNWSACKLGGGCGEGVQIRSLSCMVHSGSI SHAAGRVEDALCGEMPFQDSILKQL 1221
DB 181 SVNLGNWSACKLGGGCGEGVQIRSLSCMVHSGSI SHAAGRVEDALCGEMPFQDSILKQL 240
QY 1222 CSVPCPGCHLTENSEWSTCELTICIDGRSFETVGRQSRRTFIIQSFENQDSCPQVLETR 1281
DB 241 CSVPCPGCHLTENSEWSTCELTICIDGRSFETVGRQSRRTFIIQSFENQDSCPQVLETR 300
QY 1282 RPCTGGKCVHYTWKASLWNNERTVWCORSQGVNVTGCSQPARPAATROCIACRKPFS 1341
DB 301 RPCTGGKCVHYTWKASLWNNERTVWCORSQGVNVTGCSQPARPAATROCIACRKPFS 360
QY 1342 YCTQGGVCGCEKGYTEIMKSNGLFDYCMKVPGESEDKADVNKLSGKNRPVNSKIHDFKGV 1401
DB 361 YCTQGGVCGCEKGYTEIMKSNGLFDYCMKVPGESEDKADVNKLSGKNRPVNSKIHDFKGV 420
QY 1402 WSLQPLDPPDGRVKIWIYGVSGAFVLLIFTVSMYIACKPKPKPQRQNNRLKPLTLAYDG 1461
```

421 WSLQPLDPDGRVKIYVGVGGAFILMIFLIFTSYLVCKPKPQHQSTPPQOKELTILAYDG 480

1462 DLDW 1465

481 DLDW 484

RESULT 14  
ADB64774  
ID ADB64774 standard; protein; 933 AA.

AC ADB64774;

04-DEC-2003 (first entry)

Human protein encoded by clone NT2RP70056690.

Human; pharmaceutical; diagnostic; gene therapy; tissue regeneration;  
cell regeneration; membrane protein; signal transduction-related protein;  
transcription-related protein; osteoporosis; neurological disease;  
cancer; tumour.

Homo sapiens.

EP1308459-A2.

07-MAY-2003.

28-MAR-2002; 2002EP-00007401.

05-NOV-2001; 2001JP-00379298.

25-JAN-2002; 2002US-00350978.

(HELI-) HELIX RES INST.

(REAS-) RES ASSOC BIO TECHNOLOGY.

Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;  
Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;  
Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;

WPI; 2003-450961/43.

N-PSDB; ADB62804.

New polynucleotides and polypeptides, useful for developing a diagnostic  
marker or medicines for regulation of their expression and activity, or  
as targets of gene therapy.

Claim 1; Page; 22app; English.

The invention discloses a polynucleotide comprising a sequence selected  
from 1970 fully defined nucleotide sequences which encode novel  
polypeptides. Also claimed is a polypeptide encoded by the polynucleotide  
or its partial peptide, an antibody binding to the polypeptide or peptide  
of the polynucleotide, immunologically assaying the polypeptide or  
peptide of the polynucleotide by contacting the polypeptide or peptide  
with the antibody of the encoded protein, and observing the binding  
between the two, a transformant carrying the polynucleotide in an  
expressible manner and an antisense polynucleotide. The oligonucleotide  
is useful as a primer for synthesising the polynucleotide, or as a probe  
for detecting the polynucleotide. The polynucleotides and encoded  
proteins are useful as pharmaceutical agents and many disease-related  
genes may be included in them, for developing a diagnostic marker or  
medicines for regulation of their expression and activity, or as targets  
of gene therapy. The genes are involved in tissue and/or cell  
regeneration. Membrane proteins, signal transduction-related proteins,  
transcription-related proteins, disease-related proteins and genes  
encoding them can be used as indicators for diseases (e.g. osteoporosis,  
neurological diseases, cancer, tumours). The cDNA may be used to regulate  
the activity or expression of the encoded protein to treat diseases. The  
sequence data for this patent is not represented in the printed  
specification, but is based on sequence information supplied by the  
European Patent Office.

XX Sequence 933 AA;  
SQ

Query Match 25.8%; Score 2126; DB 7; Length 933;  
Best Local Similarity 49.6%; Pred. No. 9.7e-118;  
Matches 389; Conservative 117; Mismatches 239; Indels 40; Gaps 12;

QY 2 VRCIQKLNRTVANEICEHFALQPTTEQACLIIPCRDVCVSEFLPWNCSKGGCKLQHR 61  
DB 158 IACIQKDKDIPAEIIICEYFEPKELLEQACLIIPQDCIVSEFSAWSECKTGGSLQHR 217  
QY 62 TRAVIAPPLFGGLQCPNLTSRACDAPISCPGLBEEYTFSLKVGPPWCKRLPHUKENPS 121  
DB 218 TRHVVAPPQFGSGCPNLTFQVCQ---SSPCEAEELRYSLHVGPPWSTCMSPHSRQVQA 274  
QY 122 GRT-----VLDNFNSDSNERVTFK-HQSYKAHHSSKSWAJEIGYQYQVQVSCTR 167  
DB 275 RRRGNKEREKDRSKGVKD--PEARLIIKKRNENRQNRQENKYWDIQLQYQREVWCIN 332  
QY 168 SDGNAMLSLCLQDSFPLTVQSCIMPDKCETSWMSWSPCKTCSRSGSLLPGRFSRNV 227  
DB 333 KTGKAADLSFCQCEKLEPMTFQSCVITKECOVSEWSEWSPCKTCHDMVSPAGTRVTRTI 392  
QY 228 KHMAIGGKCEPELLEKEACIVEGELLQCPRYSWRTSEWKECOVSLLEQODPHWHTG 287  
DB 393 RQFFIGSEKCEPEFEKEPCLSQDGVVPCATYGMRTTEWTECRVDPPLLSQODKRRNQ 452  
QY 288 PVOGGGIQTRVYCAQS-----VFAAALRAKEYSRPVEKALCVGPAPLPQCLNIPGSTD 343  
DB 453 ALCGGGIOTREVYCVQANENLLSLSLTHKNKEASKPMDLKLCITGPIPTNTQLCHIPCPT 512  
QY 344 CIVSSWSAWGLCTHENCHEPQKGGFTRORHVLMESTGPA---GHCPLHVESVPCEDPM 400  
DB 513 CEVSPWSAWGPCTYENCNDQGGKGFKLRRRTTNEPTGGSGVTGNCPLHLEAIPCEPA 572  
QY 401 CYRWLASE-GICFPDGHK-CGLGHRILKAVCONDRGEDVSGSLC-----PVPPPPERKSC 453  
DB 573 CYDWKAVRLGNCEPDNGKEGPGTVQVQEVVCIINSDEVDVQLCRDALFIP-----VAC 627  
QY 454 EIPCRMDCVLSEWTEWSSCSQSCSNKNSDKQTRSRITLALAG-EGKPCPPPSQALQHR 512  
DB 628 DAPCPKDCVLSTWSTWSSCSHTCSGKTGKQIRARSILAYAGEEGGIRCPNSSALQEV 687  
QY 513 LCNDHSCMLQHWETS PMGCS EDTLVLTALNATIGWGEATCGVGIOTRRVFCVKSHVGV 572  
DB 688 SCNEHPTVYHWGTGPGWGCIEDTSVSSFNITTTWGEASCSCVGMQTRKVICRVNVGV 747  
QY 573 MTKRCPDSTRPVRPCFLPCKKDCIVTAFSEWTPCPRMQAGNATV-KQSRVRIIIQEA 631  
DB 748 GPKKCPESLSPETVRPCLLPCKKDCIVTPYSDWTSCPSSCKEGDSSIRKQSRHRVILP 807  
QY 632 ANGQOECPTLYERECEDEVSLCPVYRWKPKWSPCLLPESVWQGITGSSSEACGKLOT 691  
DB 808 ANGRCDCPTLYEKEACEAPQACQSYRWKTHKWRRCQLVWMSVQDQSPGAGCGGPGRA 867  
QY 692 RAVSCISDDNRSAMMECLKQTMGMLLVQECTVPCREDCTFTAWSKFTPCSTNCEATKS 751  
DB 868 RAITCRKQDQGGAGIHECLQVAGVPALTAQACQIPQDDQLTWSKFSKSCNCGDCGAVRT 927  
QY 752 RRRL 756  
DB 928 RKRTL 932

RESULT 15

ABR58460

ID ABR58460 standard; protein; 577 AA.

XX ABR58460;

XX ABR58460;

DT 07-JUL-2003 (first entry)

XX Human NOV47k.

DE



XX		Human; NOV; antidiabetic; anorectic; antibacterial; virucide;
KW		immunomodulator; cytostatic; nootropic; neuroprotective; dyslipidaemia;
KW		antiparkinsonian; antilipemic; gene therapy; metabolic disorder;
KW		diabetes; obesity; infection; cachexia; cancer; Parkinson's disease;
KW		neurodegenerative disorder; Alzheimer's disease; immune disorder;
KW		haematopoietic disorder.
XX		
OS	Homo sapiens.	
XX		
FN	WO2003029423-A2.	
XX		
PD	10-APR-2003.	
XX		
XX	02-OCT-2002; 2002WO-US031358.	
PF		
XX		
PR	02-OCT-2001; 2001US-0326483P.	
PR	05-OCT-2001; 2001US-0327342P.	
PR	09-OCT-2001; 2001US-0327917P.	
PR	09-OCT-2001; 2001US-0328029P.	
PR	09-OCT-2001; 2001US-0328044P.	
PR	09-OCT-2001; 2001US-0328056P.	
PR	12-OCT-2001; 2001US-0328849P.	
PR	15-OCT-2001; 2001US-0329414P.	
PR	17-OCT-2001; 2001US-0330142P.	
PR	22-OCT-2001; 2001US-0341058P.	
PR	24-OCT-2001; 2001US-0339266P.	
PR	24-OCT-2001; 2001US-0343629P.	
PR	29-OCT-2001; 2001US-0349575P.	
PR	01-NOV-2001; 2001US-0346357P.	
PR	12-APR-2002; 2002US-0371974P.	
PR	12-APR-2002; 2002US-0371980P.	
PR	17-APR-2002; 2002US-0373261P.	
PR	19-APR-2002; 2002US-0373805P.	
PR	23-APR-2002; 2002US-0374738P.	
PR	16-MAY-2002; 2002US-0381101P.	
PR	17-MAY-2002; 2002US-0381635P.	
PR	29-MAY-2002; 2002US-0383830P.	
PR	01-OCT-2002; 2002US-0026283P.	
XX		(CURA-) CURAGEN CORP.
XX		
PI	Alsobrook JP, Anderson DW, Boldog FL, Burgess CE, Catterton E;	
PI	Edinger SR, Ellerman K, Gerlach VL, Gorman L, Guo X, Ji W;	
PI	Kekuda R, Leach MD, Li L, Miller CE, Patturajan M, Rieger DK;	
PI	Rottenberg ME, Shinkels RA, Smithson C, Spytek KA, Taupier RJ;	
PI	Vernet CAM, Voss EZ, Zerkusen SD, Zhong M;	
XX		
DR	WPI; 2003-381625/36.	
DR	N-PSDB; ACC72172.	
XX		
PT	NOVX polypeptides and nucleic acids useful for diagnosing, preventing or	
PT	treating NOVX-associated disorders, e.g. diabetes, obesity, cancer or	
PT	dyslipidemia, and in chromosome mapping, tissue typing or	
PT	pharmacogenomics.	
XX		
PS	Claim 1; Page 265; 487pp; English.	
XX		
CC	The present invention relates to novel human NOV proteins and their	
CC	coding sequences (ACC72075-ACC72181 and ABR58363-ABR58469). The NOV	
CC	proteins are useful in manufacturing a medicament for treating a syndrome	
CC	associated with a human disease. The NOV proteins and coding sequences	
CC	may be used to diagnose, treat or prevent metabolic disorders such as	
CC	diabetes or obesity, infections, cachexia, cancer, neurodegenerative	
CC	disorders such as Alzheimer's disease or Parkinson's disease, immune	
CC	disorders, haematopoietic disorders and various dyslipidaemias	
XX		
SQ	Sequence 577 AA;	
Query Match	21.4%; Score 1762.5; DB 6; Length 577;	
Best Local Similarity	52.2%; Pred.No. 2.le-96;	
Matches 298; Conservative 102; Mismatches 162; Indels	9;	Gaps 4

[illegible]

Search completed: February 24, 2004, 01:02:08  
Job time : 148.011 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

un on: February 24, 2004, 00:32:36 ; Search time 51.0517 Seconds  
(without alignments)  
2760.353 Million cell updates/sec

itle: US-10-022-710-2

effect score: 8241  
equene: 1 MVRGIQLNRIVVANEICEH.....QSTPPQKPLTLAYDGLDM 1465

coring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

searched: 283366 seqs, 96191526 residues

otal number of hits satisfying chosen parameters: 283366

inimum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase :

PIR\_78:\*  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	694	8.4	2165	2 T21371	hypothetical prote
2	587.5	7.1	1444	2 T18856	angiogenesis inhib
3	429	5.2	1059	2 T22545	hypothetical prote
4	428.5	5.2	807	2 A38152	F-spondin - rat
5	405.5	4.9	1558	2 C89114	protein G37C3.6a {
6	405.5	4.9	2167	2 T34395	hypothetical prote
7	393.5	4.8	803	2 A47723	F-spondin precursor
8	357.5	4.3	1700	2 S08167	Balbani ring 3 pr
9	345.5	4.2	1074	2 JCS928	senaphorin F precu
10	345	4.2	998	2 T14764	hypothetical prote
11	340.5	4.1	788	2 T25061	hypothetical prote
12	340.5	4.1	951	2 T00260	hypothetical prote
13	335	4.1	957	2 T15976	hypothetical prote
14	335	4.1	5376	2 T42215	zonadhesin - mouse
15	333	4.0	1360	2 T33922	hypothetical prote
16	316	3.8	805	2 T34212	hypothetical prote
17	312.5	3.8	469	2 S23126	properdin precursor
18	310.5	3.8	2321	2 S78549	notch3 protein - h
19	309	3.7	2318	2 S45306	notch 3 protein -
20	305	3.7	2871	2 A55567	fibrillin I - bovi
21	302.5	3.7	1584	2 T00026	brain-specific ang
22	302	3.7	2703	1 A24420	notch protein - fr
23	298	3.6	2437	2 S42612	transmembrane prot
24	297	3.6	2307	2 A57278	fibrillin-2 precu
25	293.5	3.6	2555	2 A40043	notch protein homo
26	292.5	3.5	2918	2 A54105	fibrillin-2 precu
27	292	3.5	2531	2 A46019	notch-1 protein -
28	284	3.4	2531	2 S18188	notch protein homo
29	283	3.4	1620	2 T27283	hypothetical prote

fibrillin 1 precu  
MEGF6 protein - ra  
hypothetical prote  
immunodominant mic  
protein CTBP - mal  
properdin - mouse  
serine proteinase  
Notch homolog prot  
thrombospondin 1 p  
fibrillin-1 precu  
antigen Em100 - E1  
hypothetical prote  
Xotch protein - Af  
brain-specific ang  
hypothetical prote  
brain-specific ang

ALIGNMENTS

RESULT 1

T21371  
hypothetical protein F25H8.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T21371, T24896  
R:Gajadsty, S.  
submitted to the EMBL Data Library, February 1996  
A:Reference number: Z19413  
A:Accession: T21371  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-2165 <N1L>  
A:Cross-references: EMBL:Z69360; PIDN:CAA93287.1; GSPDB:GN000022; CESP:F25H8.3  
A:Experimental source: clone F25H8  
R:Gajadsty, S.  
submitted to the EMBL Data Library, February 1996  
A:Reference number: Z19949  
A:Accession: T24896  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-2165 <W12>  
A:Cross-references: EMBL:Z69361; PIDN:CAA93288.1; GSPDB:GN000022; CESP:F25H8.3  
A:Experimental source: clone T13H10  
C:Genetics:  
A:Gene: CESP:F25H8.3  
A:Map position: 4  
A:Introns: 31/1; 52/1; 135/2; 193/3; 216/1; 266/1; 495/2; 547/3; 584/3; 634/2; 744/1; 814

Query Match	8.4%	Score	694;	DB	2;	Length	2165;
Best Local Similarity	22.9%	Pred. No.	3.7e-36;				
Matches	351;	Conservative	133;	Mismatches	556;	Indels	492;
Gaps	84;						
Qy	18	CEHFLQPPTEQACLPDPR----	DCVSEFLPWNCSKCGKLOHTRAVIAP-PIF	71			
Db	586	CHH-----GACVRLAPESLTKIDQMGDMWGECSRTCGGVQKGLRDCDSKPRN	637				
Qy	72	GGLCQPNLTES-RACDAPISCPGEEYTFSLKVGPMWCKRLPHLKEINPNSGRTVLDPNS	130				
Db	638	GGKYCVQGRERYSCNTQ-ECFWDTPQY-----REVQCS-----EFN-	673				
Qy	131	DSNERTVFKHQSKAHHSKSWAIEIGYQTRVSCVTRSDQGNMLSL-----CLQD	181				
Db	674	--NKDIGIQGVASTNTHWPKYANVAPNERCKLYC-RLSGSAAFYLLLRKVDVGTGPDEN	730				
Qy	182	SFPLTVQSCIMPKDCE-----TSQSSWSWSPCKTCSGSLIPG-----	219				
Db	731	GGDLCVAGACMPAGCDHQLHSTLRDKCGVGGDDSSCKVVKGTFTNEQGTFGVNEWKIP	790				
Qy	220	FRSRNRNVKMAIGGGKCEPELLKEACIVEGELLQCCPRYSWRTSEWKECQVSLLEQ-	278				
Db	791	AGSANIDIRKQYNNKEDDDNYLSLA--ANGEFLLN-----GHFQVSLARQOI	837				

QY	279	--OD-----PHWHTGP-----VCGGGIQTREV---YCAQSVPAAALRA	313
Db	838	AFQDTVLEYSGSDAIIERINGTPIRSDIIYVHVLSVGSHPPDISYEYMTAAVNA-----	892
QY	314	KEYSRPVEKAL-----CVGPAP-----LPSQ-	334
Db	893	--VIRPISSLYLWRVTDTWTECDRACRGQOSQKLMLDMSTHQSHDRNCQNVLKPQA	950
QY	335	--LCNIPCTDCIVSSWSAMGLCIHENCHBPQGKKGFRTQRHVLMBSTGPAHCPLHVE	392
Db	951	TRMCNIDCSRWITEDYS-----SCSAKSGSQKRQVRVCXMGBDGQTPASEHLCD	1002
QY	393	--SVPCEDPWCY-----RWLASGEI-CFPDHKCGLGHRILLKAVCONDERGEDVSGLC-	442
Db	1003	RNSKPSDIACYIDCSRKNWYGWTSCSTCGSNGKHRR--KSYCVDDSNRRRVDESCLG	1060
QY	443	-PVPPPPEKRSCHIPCRMDCVLSWT--SWSSQSQCSENKNSDKGTQPRSTITILALAGE--	497
Db	1061	REQEATERECNIPC-----PRWVGHNSECRSCDG---GVKXHAQLDAADRET	1110
QY	498	GKGPCPPSOALQEHRLNDHSCHMOLHWETSPWGPCSEDTLVLTAINATIGWNGSATCVGI	557
Db	1111	HTSRCGPAOT-QEH--CNEHAC-TMWQFGWMSDCS-----AKCGDGV	1148
QY	558	QTRRVFCVKSHVGVMTKRCPDSTRPTVPFCLPKCKDCIVTAFSWTPCPRMQ---	613
Db	1149	QYRDANCTDHRSVLPHEHRLKWEKIIT-KPEH--RESCEPYKLGBWSQCSVEDGWS	1204
QY	614	-----AGNAT-----VKQSYRIIIOBAANGQECP-----	639
Db	1205	SRRVSCVSGNGTEVMDSLCGTASDPASHQTCNLG-TCPFWRNTDSACSVSOGIGHRR	1263
QY	640	--DTLYERBECEDVSLC-----PYVRWKPKQWSPCILVPSVWQGITGS	682
Db	1264	TTECIYREQSV-DASFQDGTKPETSQTHLLPCTSWMKPSHWSP-----S	1308
QY	683	EACGKGJQTRAUSC-----ISSDNRSAEIME-CLKQTNGMPLLVBQECTVPCRE	729
Db	1309	VTCGSGIQTBSVSCRSGEGTIUDEYFCDRNTRPLRKTKCEKQTCDCGRVLQK----	1364
QY	730	DCTFTANS--KFTPCSTNCEATKRRRLQLTGKSRRKEKCOD-----SDLYPLVET	777
Db	1365	DVPPIRWATGWTATCATCG-----NGTORLLKCRDHVDRLDEPECNHLDEKVEST	1415
QY	778	ELCPCDFPISQPYGNWSDCILLPEGRBPHRLGLYQADSKEGEGLEFRFAVACSDXNRGPV	837
Db	1416	RNGLRDCSYWMAEWEECPATCG-THVQQRNVTCVSAEDGGHFIUKDYDC-DVQKRPT	1473
QY	838	DPSPCSSGYIQEKVIPICP-FDCKLSDW--SSWGSCSSGCGIGVRIRSKWLKEKPYNGG	894
Db	1474	SARNCR-----LBPCPGEBHGSIWIIGDWSKCASCGGWRRRSVSTSSCDET	1524
QY	895	RPCPKLDLKNQVHEAVPCYSECNQ-----YSMVVEHWSSCKINNELSLRCGGGTQS	946
Db	1525	RK-PKM-----FDKCNSELCPPLTNNSWQISPWTHC-----SVSCGGGVQR	1564
QY	947	RKIRCUNTAGGGAVIDNLGNQDEI-----PBETQSC-----SL	981
Db	1565	KINCEDVLSGR-----KQDDIECSEIKPREQDCMPCCRSYHNKNTSSASMTSL	1615
QY	992	MCPNIECVMS-----BW--GLWSKCPQSCDPTHMQRTHRLLLRPSLN-----S	1021
Db	1616	SSSNSNTTSSASASLPILPFWVSWQTSAMSAKACGGRGT-KERVVECVNPSLVNTVAS	1674
QY	1022	RTCAEDSQVPC---LILNENCFOYNLTESWSTCOLSENAPCGQGVTRLILLSVCSDGK	1077
Db	1675	TEC--DQTKXPVEEVCRFKIC--PRWKTTTWSSCSV-----CBGRTARREVQCYRGKN	1726
QY	1078	PVSMDOQCOHNLKQPQMSIFCLVECV-VNOLSOW--TAWTECSQTCGHGGRMSRTRFI	1134
Db	1727	LVSDBSEC-----PKTLKLSVANCFPVACPAYRVWNVTWPSKCKDECARQOKQTRRHVC	1779

Qy	1135	IMPTQGEGRPCPTELTQEKY	-----CPVTPC-YSNVLGNWSACKLEGGDQCEGVQIRSLSC	1189
Db	1780	ISTSGKRAAPRMCLARAPTSIRECTNSCPYEWVPGDMQTC	-----SKSCGEGVQTRVRC	1836
Qy	1190	--MVHSGSIASHAAGRVEDALCGEMPFQDSTLKQKCSV-PCPGD	-----CHLT-----EWS-	1236
Db	1837	RRKINFSNIIPIFMLED	-----EPAPVKEKCELFKPKNESQICELNCPDSEFKWSP	1888
Qy	1237	EWSTCELTFC	-----IDGRSPETV	1254
Db	1889	GPWGECSKNGCGQIRRRRVKCVANDGRRVERV	1920	
RESULT 2				
T18856				
angiogenesis inhibitor homolog - Caenorhabditis elegans				
C;Species: Caenorhabditis elegans				
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000				
C;Accession: T18856; T24653				
R;McMurray, A.				
submitted to the EMBL Data Library, July 1995				
A;Reference number: Z19031				
A;Accession: T18856				
A;Status: preliminary; translated from GB/EMBL/DDBU				
A;Molecule type: DNA				
A;Residues: 1-1444 <W1>				
A;Cross-references: EMBL:Z50004; PIDN:CAA90293.1; GSPDB:GN00028; CESP:C02B4.1				
A;Experimental source: clone C02B4				
R;McMurray, A.				
submitted to the EMBL Data Library, July 1995				
A;Reference number: Z19917				
A;Accession: T24653				
A;Status: preliminary; translated from GB/EMBL/DDBU				
A;Molecule type: DNA				
A;Residues: 1-1444 <W12>				
A;Cross-references: EMBL:Z50006; PIDN:CAA90302.1; GSPDB:GN00028; CESP:C02B4.1				
A;Experimental source: clone T07C5				
C;Genetics:				
A;Gene: CESP:C02B4.1				
A;Map position: X				
A;Introns: 25/3; 70/3; 96/3; 139/3; 187/1; 234/2; 282/3; 376/2; 422/2; 478/3; 5				
Query Match 7.1%; Score 587.5; DB 2; Length 1444;				
Best Local Similarity 23.4%; Pred. No. 1.5e-29;				
Matches 315; Conservative 155; Mismatches 460; Indels 419; Gaps 86;				
Qy	12	VVANEI-----CSHFALOPTTEQAQLPCRDGVSEF	-----LPNSNCS-----	51
Db	368	VLAHEHNNGMVHDGVQNCNKG	-----CLMSAVNGAGKTTSDCVRFENAFLLQ	420
Qy	52	---KGCGKKLQHRTRVIAAPFLFGLQCP--NLTEGRACDAPISCPGLGEBEYTFSLKVP	106	
Db	421	LDSESGRNCRLDASPLGIISTNHLSDLRLPQRFADQCC	-----SYFWGR	465
Qy	107	WSKCRILPHLKEINPSGRVTLDVFNDSNERYTFPHQSYKAHHHKSMALEIGYQTRQVSCT	166	
Db	466	DYKVEIPNGKAMDIDICILKCGNSGST--ISTAHPALEG	-----SWC-----	505
Qy	167	RSQGNAML--SLCLQDSFPLTVQSCIMPDCETSDWSSSPSCSKTCR	-----SGSILP	218
Db	506	---GANKWCHKQCTHTWFTGLT-----VPID--GEWSENGGAEGKCPIQCCAVSGSITV	555	
Qy	219	GFRSRNRNVKHAIGGKECPLELLEKAC---IVEGELLQCPRYSWRTSEWKECVSL	274	
Db	556	QGQHRD-CVNPAPNNGGKTC-----EGANIRGIVCGATSSNCLGFTREFFGNKICS-SI	607	
Qy	275	LLEQQDPHHWHTGVPCGGGQITREVYCAQSVPAALAAKAEVSRPEKALCVGAPLP SQ	334	
Db	608	KYDPHKDQQTGEGFHSHTQPCRVWC--HLIGSELIRNG	-----	646
Qy	335	LCNTPCSTDCIVSSWSAWGLCIHENCHEPOGKGFRTQRHVLWMESTGPAGHCPHLVESV	394	
Db	647	--OPFDGTPCGFDAYCVGGGCLALSC-----DNKALVEQ---PEDCPRI---	685	



```
QY 1091 -----KPO-----RMSIPCLVBCVNCQLSGWTAMTECSQTGCGG 1126
Db 746 TLNFKYRDPASRVLDLAPKQIESTKXNRQVQVSKEDVLRQAS-----VLHKM 794
QY 1127 RMGRTRFIIMPTQEG-----RCPPTLTQKTCPTVPCYSWVLGNWSACKLEGGDCG-B 1180
Db 795 NVSLIEALLTAPNDKAREQLRKYGNELVAR-----WDIGHWSECRQK--TCHVA 842
QY 1181 GVOIRSLSCMV-HSGSIASHAAGRYEDALCGEMPFQDSILKQLCSV-----PC-PGDCHL 1232
Db 843 GYQARGISCKVTFGEIRN-----VDNSIC-----ESLASVRPPTPRCHREDCPR 888
QY 1233 TEWSEWSTC-ELTCIDGRSPETVGRQSRRTFIIQSFENQDSQPQ--VLETRPCTGGKC 1289
Db 889 WEASOWSECSQRCS-----SMLAQKRNVTCRFTNGTSVDIQHCDITNRPATTMD 941
QY 1290 YHYT-----WKASLWNNRITVWCORSQGVNVTG-----CSFQAR 1325
Db 942 PNQCKAEWRTSDNGS-----CSSECG-----TGGVQLRLSCVWISGRPAGRNCEQMR 992
QY 1326 PAATROCI-----PACRKPFSYCTQ-GGVGCGCKGYTEIMK 1360
Db 993 PHSARACVADEPLPCMPPTASALYQRDASQDOSRPFCDIHK 1033

RESULT 4
A38152
C:Species: Rattus norvegicus (Norway rat)
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 17-Nov-2000
C:Accession: A38152
R:Klar, A.; Baldaasare, M.; Jessell, T.M.
Cell 69, 95-110, 1992
A:Title: F-spondin: a gene expressed at high levels in the floor plate encodes a secreted
A:Reference number: A38152; MUID:92208952; PMID:1555244
A:Accession: A38152
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-807 <KLA>
A:Cross-references: GB:M88469; NID:g204176; PID:AAA41174.1; PID:g204177
A:Experimental source: embryo floor plate
C:Superfamily: F-spondin; thrombospondin type 1 repeat homology
F:441-495/Domain: thrombospondin type 1 repeat homology <THR2>
F:500-555/Domain: thrombospondin type 1 repeat homology <THR3>
F:557-611/Domain: thrombospondin type 1 repeat homology <THR4>
F:613-666/Domain: thrombospondin type 1 repeat homology <THR1>
F:667-721/Domain: thrombospondin type 1 repeat homology <THR5>
F:753-807/Domain: thrombospondin type 1 repeat homology <THR6>

Query Match 5.2%; Score 428.5; DB 2; Length 807;
Best Local Similarity 28.8%; Pred. No. 1.2e-19;
Matches 133; Conservative 64; Mismatches 158; Indels 107; Gaps 27;

QY 849 QEKVIFCPFDCKLSDWSSWGSCSS--CGIGVIRSRKWLKEKPYNGRPPCLDLKNQVH 907
Db 432 EEKDDDTPETCIYSNWSPNSACSSSTCEKGRVQRMLKQAQ-LDLVSPCP-----DTQ 484
QY 908 EAVPCY-----SECNQYSVWVEHSSCKINNELRSRLCGGTSRKIRCVNTADGE 958
Db 485 DFQPCMGFGCSDEBGTCTMSEWIT--WSPC-----SVSCGNGMRSRE-RYKTOPPED 534
QY 959 GGAVDSNLQNDIEP-PETQCSL--MCPNECVSEWGLWSKCPQCDPHTMQRTRHL 1014
Db 535 G-----SVC-----MLPTEKETCTVNECSFSSCLVTEWGEWDDCSATCGMKKGRWVK 586
QY 1015 LRPSLNRCTAEDSQVQCLLNENCFQVNL-----TEWSTCQLSENAPCGQVRT--RL 1069
Db 587 MSPADGSMCKAETSAQKCMWPE-CHTIPCLLSPWSESDCSVT-----CGKMRTRQRL 641
QY 1070 SCVCSGDKFVMDQEOHNLKPMRMSIPCLVECVNCSLQSGWTAMTECSQTGCHGRMS 1129
Db 642 KSLAELG-----DQNE-DLEQAEKMLP--ECPIQCELSWSEWSECNKSCOK-GHMI 690
```

```
QY 1130 RTRILMETGEGRPCTELTQKTCPTVPCYSWVLGNWSACKLEGGDCGEGVQIRSLSC 1189
Db 691 RTRIQMEPQGGAPCP-ETVQRKKCRARKC-----LRSPSIQKLWRERARE----- 736
QY 1190 MVHSGSISHAAGRYEDALCGEMPFQDSILKQLCSVPCPGDCHLTWSEWSTCBLTCTIDG- 1248
Db 737 -----SRRSEQLRESDGEQ-----PPG-CRMRPMTAWSECTKLCGGGI 774
QY 1249 -RSPETVGRQSRRTFIIQSFENQDSQPQVLETRPCTGGKC 1289
Db 775 QERYNTVKKRFKSSQF-----TSCKDK-KETIRACNVHPC 807

RESULT 5
C89114
Protein C37C3.6a [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C:Accession: C89114
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
A:Reference number: A75000; MUID:95069613; PMID:9851915
A>Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/projects/C_ele
A>Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: C89114
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1558 <SFO>
A:Cross-references: GB:chr_V; PIDN:AAC25867.1; PID:g3294501; GSPDB:GN00023; CBSP:C37C3.6
C:Genetics:
A:Gene: C37C3.6a
A:Map position: 5

Query Match 4.9%; Score 405.5; DB 2; Length 1558;
Best Local Similarity 20.3%; Pred. No. 6.6e-18;
Matches 322; Conservative 136; Mismatches 502; Indels 629; Gaps 86;

QY 46 PW---SNGSKGCKKLQHRTRAVIAPLFGGLQCPNLTESACDAP-----ISCPLG 94
Db 81 PWPVNECSRSCG-----GGVQ-----LEKQCSGDCGTGASVRYISCNLN 120
QY 95 EEEYFSLKVPWSKRLPHLKEINPSGRTVLDNSDSNERVTEKHOSYKAHHSK----- 150
Db 121 ACESGTDFAQCCK-----FNDALDGNHYKHTFPYKGNKCELVK 162
QY 151 -----SWAIEIGYQTRQVYCTSRSDGNAMLSLQDSPLTVQSCIMPDKCETQWS 202
Db 163 PBSGNFYFKWADKV-----VDGTKCDSK-----SNDICVDGECPLFVGCDGKLGS 206
QY 203 S--WSPCSTKCRSGSLLPGRSR--SRNVKMAIGGKCEPPELKEACIVEGELLQOCP 258
Db 207 SUKPKCKCKDGDGSGTCKTIEGRFDERNLS-----PGVHDIILKPEGATNI----- 252
QY 259 RYSMTSEWSEKQVSLLEQDDPHWHVTGPGVGGGIOTREVIYCAQSVPAALAAKAYSR 318
Db 253 ----KIQEARSTNNALKNKGDHFLYNG---NGLIQ----- 282
QY 319 PYEKALCVG-----PAPLPSQLCNIPCSITDCIVSSWSANGLCIHENCHBPQKKG 368
Db 283 -YEKEVEYGGTIFVYDDAEPETLSAQ---GFLSELTV-----LIFRQSGRDTAIKYE 332
QY 369 FRTRQRHVLMESTGPAGCHPLHVESVPCEDPMCY-----RWLASEGICFPDHGKGLGHR 423
Db 333 F-----SIPLEEEVDYMYKFDNW-----PCSVSCGKGQV 362
QY 424 ILKAVQONDRGED-VSGSLCP---VPPPPERKSEIPCRMDCVLSEW--TEWSSCSQSCS 477
Db 363 TRNLYCIDGNKGRVEDDLCEBNNAKPEFKSCB---TVDCE-AEWFTGDWESCSSTCG 418
QY 478 KNNSGKQTR-----SRTILALAGEGCKPCPPSQALQEHRLCNHDHSCMQLHWETS 527
```

b	419	DQ----	GOQYRVVCHQVANGRAVTVEDGNCTVPRPVK-----	QTCNRFACPE--WQAG	468		
y	528	PWGPCSDTLVTALNATI	GNGEATCGVIGIQRTRVFCVKSHVGQ-----	VMTKRCP--DST	581		
b	469	PWSACE-----	KCGDAFOYRSVTCRSEKEGEBKLLAADACPADEQE	511			
y	582	RPEVTRPCFL--PCKKOC	IVTASEWTPCPRMC-----	QAGNATVQSYRIIIOE--	634		
b	512	KFDVERTCNJGPCEGLT	FTVT--GEMNLCTR--	CNDTEETREVTCKDSQGRAYPFEKCLVDN	568		
y	635	GQECFDLYIERECEDV	SLCPVYRWKPKQWSPCILVPBSVMQGITGSGEACGKLOTRAV	694			
b	569	STEIPT--	DTRSCATOPPCE--	YEMTVSWSXC-----	TTECGHGKHTRV	610	
y	695	SCISDDNRSAEMMECL	KQTNMGPLLVQECTVPCREDCTFTAWSKFTPCSTNCEATKGRRR	754			
b	611	ICAIHQNGGLEVD--	-----	EGHCOAEKPE--	634		
y	755	QLTKGR--	KKEKQCDSDLYPLVETELCPDEFISQPYGNMSDCILPEGRREPHRGLRVQ	812			
b	635	--	GKTNCTNEEKTGT--WYT-----	SSWSEC-----	657		
y	813	ADSEKCEGLRFR	AVACSDKNGRVPDPSFCSSSGYIQEK-----	CVIP-----	CP--857		
b	658	--TAE	CGGQDRVAVCLNYDKKPV--PEWCDZAVKPSBKQCNQDNDCTCVDSEFGCCPD	714			
y	858	-----	FDCKLSDMSWSSGSSSCGIGVIRBSKWLK--	EKPYNGRPPCPKLDKN	904		
b	715	NSTFATGEFNG	CSNSETFEGCDAUNVTATGPNMSKGCERFVSPLN-----	LEADVAN	769		
y	905	QVHEAVPCYBECNQ	YSVVVHEHNSCKINNELSLRCGGTOSRKIRCVNTA-----	DG	957		
b	770	ADAEASGAPEL--	-----	CSVTNE-----	NGEADVECATIAPIALLGDG	809	
y	958	EGGAVDSNLCNDEI	PIPETQSCSLMPCNFCVMSWGLMSKCPQSCDPTMQORRTHLLRP	1017			
b	810	ELINDYDASN-----	ETIHCS-----	KTERGC--CP-----	834		
y	1018	SLNSRTCAEDSQVQPC	LLNENCQFOFY--NLTEWSTCQ-----	LSENA--PCQQGVTR	1067		
b	835	--	DWYTAASGK-----	NEGCFSTLGCNETQGCCHDDVTLARGANLEGCEPSCAA	886		
y	1068	LLSCVCSDDGPVSM--	DOCEQHNLKQRMISICLVECVNQLSGTAWTECSOTG	1123			
b	887	SLYGCCDRKRTIA	FGPHYSCE-----	RSFFPC-----	914		
y	1124	HGGRMSRTRFI	IMPTEQEGRPCTELTKETCPVTPCYSWVLGNWSACKLEGGDCGEGVQ	1183			
b	915	-----	ELSDFGCCPD-----	GETAALGKNGTCCGNCIL	942		
y	1184	IRLSJC-----	MVHSGSISHAAGRVEDALCGEMPTQDSILKQLCSVPQDCHLTSEW	1238			
b	943	TTKFGCCPDGKTTAK	SGHNBGGC--CEFAQYGCPCDGKSVAKGAGYGCPSCAQSQFG--	999			
y	1239	STCELTCIDGRSFET	VGQRSRSTFIIOGFENODCPQVLETRPCGCKYHYTWKASL	1298			
b	1000	-----	CCPDGK-----	TRARG-----	ENKEGCPQYTRYGCCPDG--	1029	
y	1299	WNNNERTVMQORS	DGVN-----	VTGGGSPQARPAAI-----	ROCIACRKPX--	1340	
b	1030	-----	ETTALGRNDGDCNCRYAXHGCCPDGETKALGPDGAGCFPTTTPFLMGGTVA	PHK	1085		
y	1341	--SYCTQ--	GGVCGCEKY-----	TEIMKSNGLDYCKMVPGBSDKADVNL--	1384		
b	1086	IAACNTOESGT	VGCA--GYKLAWHYDTTEGRCNQFYGGC--	GGNDNNFASQDV	CET	1139	
y	1385	-----	SGKRNRPVNSKIHDI	FKGWSLOP--	LDPDGR--	VKIWYGVSGAFLIMIFLIFT	1434
b	1140	ICVFPFGKRCYL	PRVDGRLPCDLOFRIYYDHSKKGCVAFWRRC	LGNA-----	NNFN	1193	
y	1435	SY-----	LVCCKPXPHQ-----	TPPQ	1452		
b	1194	SFECSMFC	KDVGVYDAPTTAA	PPPPPQ	1222		

## RESULT 6

T34395  
Hypothetical protein C37C3.6b - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 01-Dec-2000  
C:Accession: T34395; T34394  
R:Geisel, C.; Bradshaw, H.  
Submitted to the EMBL Data Library, July 1996  
A:Description: The sequence of *C. elegans* cosmid C37C3.  
A:Reference number: Z21518  
A:Accession: T34395  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2167 <GE1>  
A:Cross-references: EMBL:U64857; PIDN:AAC25868.1; GSPDB:GN00023; CESP:C37C3.6b  
A:Experimental source: strain Bristol N2; clone C37C3  
A:Accession: T34394  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1555 'SKP' <GE2>  
A:Cross-references: EMBL:U64857; PIDN:AAC25867.1; GSPDB:GN00023; CESP:C37C3.6a  
A:Experimental source: strain Bristol N2; clone C37C3.  
C:Genetics:  
A:Gene: CESP:C37C3.6b; CESP:C37C3.6a  
A:Map position: 5  
A:Introns: 32/3; 104/2; 156/2; 207/1; 459/2; 536/3; 577/2; 1105/3; 1367/1; 1438/1; 1556/

Query Match 4.9%; Score 405.5; DB 2; Length 2167;  
Best Local Similarity 20.3%; Pred. No. 9.3e-18;  
Matches 322; Conservative 136; Mismatches 502; Indels 629; Gaps 86;

46	PW---	SNCSKGCKLOHTRAVIAPPLFGGLOCPNLTSRACDAP-----	ISCPIG	94	
	:	:	:		
81	PWPENECRSQC-----	GGVQ-----	LEKQCSGDCGTGASVYISCNLN	120	
	:	:	:		
95	EEBYTSLXGVPWSKCLRPLHLKEINPSGRVTLDFNSDSNERVTFKHQYKAHHSK---	150			
	:	:	:		
121	ACBSGTDFRAEQCSK-----	FDEALDGNVHKVTPYKGNKNCLEVK	162		
	:	:	:		
151	-----	SWAIBIGYQTRQVSCSTRSGONAMLSLCLQDSFPPLTVQSCINPKDCETSQWS	202		
	:	:	:		
163	PESGNFYWKADKV-----	VDGTKCDSK-----	SNDICVDGECLPVGCDGKLG	206	
	:	:	:		
203	S--	WSPCSKTCRSGSLLPGRSR--	SRNVKHMIAIGGKECPELLEKEACIVEGELLQOCP	258	
	:	:	:		
207	SLKFDKCGKCDGSGCTKIEGRFDERNLS---	PGYHDIILKPEGATNI-----	252		
	:	:	:		
259	RYSWRTSEMKECQVSLLLQQDPHMHVTPVCGGGJQTRREVYCAQSVAAAAALRAKEVSR	318			
	:	:	:		
253	-----	KIQEARKSTNNALKNGSDHFYLNG--	282		
	:	:	:		
319	PVEKALCVG-----	PAPLPSOLCNPICSTDCIVSSWSAWGLCIHENCHEPGQKKG	368		
	:	:	:		
283	-	VEKEVEVGITFYVDDAEPTLSAQ--	GPLSEELTVA-----	LLFRKGRSDTAIKYE	332
	:	:	:		
369	FRTRQRHVLWESTGPGAGCHPLVESVPCDDPMCY-----	RWLASEGICPPDHQKCGLGH	423		
	:	:	:		
333	F-----	SIPLEEVDMYKFDNWT-----	PCSVSCGKG	362	
	:	:	:		
424	ILKAVCONDEGED-VSGSLCP--	VPPPERKSKCEIPCWMDCVJLSEW--	TEWSSCSOSCS	477	
	:	:	:		
363	TRNLYCIDGKNKGRVEDDICEENNATKPFESKCE---	TVDCB-ALWFTGDMWSCSSTCG	418		
	:	:	:		
478	NKNSDGKQTR-----	SRITILAGEGGKPCPPFSQALQEBHRLCNHSCNQMLHWETS	527		
	:	:	:		
419	DQ--	CGQYRVVYCHQVFANGREVTVEDGNCVTERPVK-----	QTCNRPACPE--	WQAG	468
	:	:	:		
528	PWGPGCEDTLTVTALNATIGNWGEATCGVGIQTRRVFCVKSHVQO-----	VMTKRCP--	DST	581	
	:	:	:		
469	PWSACSE-----	KCGDAFQVRSVTCRBEKEEGEGLLAADACPADEQE	511		
	:	:	:		





Residues: 1-1700 <PAU>  
Cross-references: GB:X52263; NID:G7057; PIDN:CAA36506.1; PID:G7058

Genetics:  
Gene: B33

Map position: 4  
Superfamily: unassigned Balbiani ring proteins

Query Match 4.3%; Score 357.5; DB 2; Length 1700;  
Best Local Similarity 19.6%; Pred. No. 8.3e-15;  
Matches 303; Conservative 146; Mismatches 595; Indels 499; Gaps 89;

28 EQACLPICPR--DCVYSEFLPKNSCKGKGLKHRTAVIAPPLFGGLQCNLTESRAC 85  
127 EKSCACVFNADKCTAPQVNNKDTCCCGFVNQEPADGCTKPLIWDKVC-----RC 179  
86 DAFISCPGLGEEYTFSLKVGPMW-----DWSDDSCSCCKGDKGCKGKIWKNCNRCICP--TAEP 228  
180 ECPFLKXDCGNR-----DWSDDSCSCCKGDKGCKGKIWKNCNRCICP--TAEP 228  
121 SGRVLDNFSDNE-----RVTFKHQSYKAHHHSKW-----AIEIGVQTRQVSCTRSDQON 172  
229 AGGCSAPLWDDDKCSCACPAKMEKKEKCVESGKIWNPTCEG--CAQLNCPDNKKAN 286  
173 AMLSCLQLDSFPLTVQSCIMPDCETSSWSNPSCS-----KTCRSGLLPQFRSR 223  
287 KETQC-----EKEVKKNGGVFCXKSCVCPGGDKXTCTAPQVYDG----- 332  
224 SRNVKMAITGGKECPPELLEKEACTIVEGELLQOCPR--YSWRTSEWK--EQVSLLEQDDP 281  
333 -----VACSCSPVNMQKPA-----DGPFRQKWDKEECRCECPVR----- 368  
282 H-----NHVT-----GPVCGGGIOTREVVYCAQSVPAALRAKEVRPVEKALC 325  
369 HDCKNGKVDETICQICPRDAPVCTAGKER-----CGES-----CECKINREPKEG-C 417  
326 VGPAPLPOLNIPCTDCIVSSWSAWG-----LCIHENCHEPOGKGFRTORHVLME 379  
418 AKPLVNNENTCKVCVPADKQMSPGGSGSKFNKLTQCECEQDSASKGLKRW----- 471  
380 STGPAGCHPLVSEVCEDEPMCYRWLASGIC-----FPDHGKGLGHRILKAVQNDRG 434  
472 -----ADTCRKE-----QPGMPPEG-CGKQTWISDKCKECSPTI-----TCQAPQI 513  
435 EDVS--GSLCPVPPPPERKSCIEPCRMDCVLSEWTESSCSQSCSN--KNSDGKOT-RSR 489  
514 LDLTCECKCPVNLQAQEKCKSP-----RGWTD-SKLCCESTTPATCGKGTWGE 565  
490 TIALAGEGKCPPPSOALQ-----EHLRNDHSCMLHWETSP--WG-----PCESTDL 537  
566 ACQCICPGDKNCKGNKFFDKPSCECKCKNNPTC-----TSPQVWDADDCECKCPKQK 619  
538 VTALNATIG--WNGEATCGVIGIOTRVRVFCVKSHVGOVMTKRCPSDRPRPVRPCLPC-- 593  
620 KPGGGDGGGOKM--DRVSCGCPVPAPDCTNGGIYNT-----CAGGCGI 664  
594 -KDCIVTAFSEPTPCRCQAGNATVKOSRYRIIIEAANGQBCPDPLYERECEVDS 652  
665 DKPSCPKQIYNNKTCDCCEPENG-----MKEPVGG--CGAKTWLDDBCQ--- 706  
653 LCPVYRWKQKNSFCILVPESVWQGITGSEACGKLQTRAV----- 694  
707 -----CDVCPKPGKGTGAKQKCDKTCCKCKEMPTGCGENKKNKWCDET 752  
695 -----SCISDDNRSAMMECL-----KQTNGMPLLVQECTVPCREDCTFTAWSKTTP 741  
753 CDCVCPQKNTCIAPKWDATCSICVNPFPKNSPQVLKDTCCCGQNVKSKCAQKRIE 812  
742 CSTNCEATSKRRRLTGKSKKKEKQDSLDYPLVETELPCDEFTSQPYGNWSDCILLEG 801  
813 NTICDCACP-----NKKQCKX-----APLVWS-----DEF-----CDCVCPN- 842  
802 RREPHRGLRVQADSKCEGLRFRVACS--DKNGRPVDPSPFC--SSSGYIOEKCVIPCP-- 857

843 -----SASMKITCLSPKWNKVTCTCDCN--PPKPDCCPGTQKWMDDKCKGCPNA 890  
858 -FDC-----KLSDSSWGSCSSSCGIG--VRIKMLKEPYNG-----GRPCPKLDLKNQ 905  
891 QTDGAGKKFNDF-----TSCGGPGLDCTGNTKWSAETCTCGGDVNRNG--NLKNF 944  
906 VHEAVPCYSECNQYVYVHEHWSKNNELRSIRCGGTQOSRKIRCVNTADGEGGAVDSN 965  
945 NDNL--CQCECKXN--KQEMANCKSPRTWYDTC-----KVCCKN-----ADSD 984  
966 LCNODEIPPTQSCSLMCPNECVNSELWSKCFQ-----SCD-----PHTM 1007  
985 DCPVQPLWLDQ--CKCCPASAQNT-----CPANKRFIEKSCSECKSPMPSPIPQK 1036  
1008 QRRTHLLRFLSLRSRTCAE-----DSQVQPL-----LNENCFQFYNLTEWSTQOLSENA 1059  
1037 KWNEDKCVVECANVKTCEGQWRWCDNCK--CICPQVNTKCDKQKFIKSCCECETQTQ 1095  
1060 CGQVTRLL--SCVSD-----GKPV--SMDQEQHNLEK-----QRMSIPCLVECVN 1106  
1096 CKDGRFWSNLECGCLDDKXCPGKQVDFKNTCQCKCPN--QKPGDTCGNGKDFCPLDCSK 1154  
1107 CQ-----LSGMTAMTECSQTCHGGRMSRTRFIIMPTQEGRGPCTELTQKTCPTVP 1159  
1155 CKNFKPANGCTGYQWNEEKQC-----ECPKD-KPKQCP----- 1189  
1160 CYSWLG--NWS-----ACKLEGDCGEGVQVRSLSVMVHSGSISHAAGRVEDALCGEM 1211  
1190 -----GGQDWNHQCQCGCTPAFTCSNNQKYSNVSC-----SCGCPKPKNGCPGNQ 1238  
1212 PFQDSILKOLC--SVPCPGDCHLTENSEWSTCLTCTIDG-----RSFET 1253  
1239 IWCNTRCVCPRKMEPADNCKTKWNDENCCVCPGCPGEGCKGKWKNNANTSCCEC 1298  
1254 VGRSRRFTIIISFENQDSCPOQVLETRCTGCKVHYTWKASLWNNNRTTYWCQRSDG 1313  
1299 PADKAKPASCQDKKSWNDSCQCKSKMPC--GCPNPQW-----NEKDCCKCS-- 1348  
1314 VNTGGC--SQAPAAIROC-IPA--CRKPFYCYTGGVCGC 1351  
1349 --ATGNCPPAGTWNSTQCSQCPATGKTGCTGAQVWCSKACKVC 1389

# RESULT 9

JC5928

semaphorin F precursor - human

C/Species: Homo sapiens (man)

C/Date: 10-Apr-1998 #sequence\_revision 08-May-1998 #text\_change 17-Nov-2000

C/Accession: JC5928

R/Simmons, A.D.; Poeschel, A.W.; McPherson, J.D.; Overhauser, J.; Lovett, M.

Biochem. Biophys. Res. Commun. 242, 685-691, 1998

A/Title: Molecular cloning and mapping of human semaphorin F from the Cri-du-chat candid

A/Reference number: JC5928; MUID:98125554; PMID:9464278

A/Accession: JC5928

A/Status: nucleic acid sequence not shown

A/Molecule type: mRNA

A/Residues: 1-1074 <SIM>

A/Cross-references: GB:U52840; NID:G2772583; PIDN:AAC09473.1; PID:G2772584

A/Experimental source: brain

C/Genetics:

A/Gene: semaf

C/Superfamily: human semaphorin F; thrombospondin type 1 repeat homology

F.1-20/Domain: signal sequence #status predicted <SIG>

F.50-533/Domain: semaphorin #status predicted <SEM>

F.840-896/Domain: thrombospondin type 1 repeat homology <THR3>

F.971-993/Domain: transmembrane #status predicted <TM>

Query Match 4.2%; Score 345.5; DB 2; Length 1074;

Best Local Similarity 22.6%; Pred. No. 3e-14;

Matches 141; Conservative 59; Mismatches 168; Indels 255; Gaps 30;

794 SDCTL-----PGRPREPHRLVQADSKCEGLRFRV----- 827

```

Db 441 SSCLEIEIEFPERRRPIRSLQILHSQSVLFVGLREHVVKVILKRCQCFYRTSTCIGAQ 500
Qy 828 -----ACSDKN-----GRPV 837
Db 501 DPYCGWDVWVKCTSLBESLSMTQEQSISACPPRNLITVDHGFVMSWPPTCTHTDGSV 560
Qy 838 DPFSCSSGYIOEKVPCP-----FDCK-----LSDWSSWGSCSSSCGIGV 879
Db 561 GSCLCRT-----RSCDSPAPQCGWQCEGFCMEIANGSRNGTPTWTSWPCSTTCGIGF 615
Qy 880 RIRSKWLKE-KPYNGSRPCPKLDLKNQVHEAVPCYSECNOYSWVVEH-----WSSCK 930
Db 616 QVRORSCNPTFRGGRCVGVQNRERY-----CNEHLLCPFMFTWGTGPHERC- 665
Qy 931 INNELSLRCCGGTQSKIRKVCNTADGEGAVDSNLNQNDEIPEPTQCSL-WCPNECVN 989
Db 666 -----TAQCGGIGQARRICENGPDGAG-----CN-----VEYQSCNTNCPKELKT 707
Qy 990 SEWGLMSKCPSCDPHTMQETRH-----LL-----RPSLSNRTCAEDSQVOPCLLN 1036
Db 708 TPWTPTIPVNISDNGDHYEQFRYTKARLADPNLLEVGQRQIEMRYCSDG-TSGCSTD 766
Qy 1037 ENCFQF-----QYNL-----TEWSTCOLSENAPCGQGVTRLLSCVCSGKP-VSM 1081
Db 767 GLSGDFLRAGRYSAHTVNGAWSMTWSQC-----SRDCSRGIRNR--KRVCNNPEPKYGG 820
Qy 1082 DQCQHNLKPKFORMSIPCLVFCVNVNQLSGHTANTSCOTCGHGRMSRTRFIIMPTQGE 1141
Db 821 MPCLGPSLEYQECNTLPCPDVGWSC-----WSPWTKSATCG-GGHYMRTRSCNP- 871
Qy 1142 GRPCPTELTQEKTPVTPCYSWMLGNWSACKLEGDCGEGVQIRSLSCMVHSGSISHAAG 1201
Db 872 -----APAY-----GGDICLGL-----883
Qy 1202 RVEALCEMFPQDSILKQLCSVPCPGDCHUTWSEWSTBELTICIDGRSPETVGRQSR 1261
Db 884 HTEALCN-----TQPCPES--WSEWSDWSECEAS-----GVQVRAR 918
Qy 1262 TFIQSFENQDSCPOVLETRPC 1284
Db 919 QCIL-LFPNGSQCSGNTTESRRC 940

RESULT 10
T14764
hypothetical protein DKFp434H204.1 - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T14764
R:Wambutt, R.; Heubner, D.; Newes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, August 1999
A:Reference number: Z18181
A:Accession: T14764
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-898 (wam>)
A:Cross-references: EMBL:AL110226
A:Experimental source: adult testis; clone DKFp434H204
C:Genetics:
A>Note: DKFp434H204.1

Query Match 4.2%; Score 345; DB 2; Length 898;
Best Local Similarity 20.1%; Pred. No. 2.7e-14;
Matches 228; Conservative 102; Mismatches 363; Indels 444; Gaps 58;

Qy 229 HMAIGGKECPLELLEKEAIVSGELLQQCFRYSWRTSEWKECOVSLLELQQDPHHHTGP 288
Db 21 HREAGHDEVPP-----PVSWHYGPWTKCTVT-----48

Qy 289 VCGGGTQTRVYCAQSPVAAAALRAKEYSRPVEKALCVGPAPLPOLNIPCTDVISS 348
Db 49 -CGRGVQRQNVYCLE-----RQAGPVDE-----70

```

RESULT 11  
T25061  
hypothetical protein T21B6.3 - Caenorhabditis elegans

```

Qy 349 WSAWGLCIHENCHBPQKKGRTRQRHVLMSTGPGACHILVESVPCEDPMCYRWLASE 408
Db 71 -----EHC-DFLGRPDDQQRK-----C-----SEQPCPARWAGE 99
Qy 409 -GICFPDGHKGLGHRILKAVQNDRGEDVSGSLCP-----VPPPPERKSC--BIPCRMD 460
Db 100 WOLCSSSGCPGLSRRAV--LCIRSVGLDEQSALEPPACEHLPRPTTETPCNRHVPCP-- 155
Qy 461 CVLSEWT--EWSSCSQSCSNKNSKQTRSTIILALAGEGKPCPPSQALQEHRLCNDHS 518
Db 156 ---ATWAGVNNQSQSVTC-----GEGTQRENVLIC-TNDTGVPCDEAQQPAVEVTCSLPL 205
Qy 519 CQLLWNETSPKGPCSEDTLVTALNATIGWNGEATCGVGIOIRRVF-----CVKSH----- 568
Db 206 C---RW---PLG-----TLGPEGs-----GGSSSHLFFNEADFFPHILARPS 243
Qy 569 -----VGQWTKRCPDSTRPETVRPCL-----PCKK-DCIV 599
Db 244 PASSPKPTMGNAIEEAPELDLP--GPVFVDFFYDYNFINFHEDLSYGFSEEPDLDL 300
Qy 600 TAFSEWTPCPKMCQAGNATVKQSYRIIIOEAANGQCEPDTLYEERECEOVSLCPVYRW 659
Db 301 AGTGDRTPPHSRPAAPST-----GSPVPATEPPAAKEGV-LGP---W 340
Qy 660 KPQKW-----SPCILVPESVWQGITGSSEACGK 687
Db 341 SPSPWPSQAGRSPPPSQTGPNPLINFLPEEDTPIGADLGLPLSLSWRVSTD----- 394
Qy 688 GLQTRAVS-----CISDDNRSAEMECIKQTN-----GMPLLVQECTVPCRE 729
Db 395 GLQTPATPESQNDFFVQKDSQSLPPPWDRDNEVFKDDEBPFGKRGAPHL-----PPRP 448
Qy 730 DCTFTAWSKFTPCSNCEATKSRRLQTKSRKKKCODSLYPLVETELCP---CDEFI 786
Db 449 SSTLPLSPVG--STHSFSPDVAELTGTIVANEPALGGLGP-VDSLELWFTVGVASIL 505
Qy 787 SOPYGNWSDCILPEGRREPHRGLRYQADSKCEGLRFRVAVACSKNGRPVD-----PS 840
Db 506 PPPIAP-----LPE-----MKVRDSLSLEPGTP-SPPAPGPGSWDLQTVAWGTFLTPT 551
Qy 841 FCSSSGYIOEKVIP-----CPFDCKLSDWSSWGSCSSSCGIGVIRSKWLKE 888
Db 552 TLTGLGHMPEPALNPGKQPESLSFEVPLSRLLSTPAWDSPANSHRPV---ETQPLAP 608
Qy 889 KPYNGRPPCKLDLKNQVHEAVPCYSECNOYSWVVEHWSCKINNELSLRCGGTQSRK 948
Db 609 SLAAGAPPADPLVVRNA-----SWQAGNWSEC-----STTCGLGAVWRP 647
Qy 949 IRCVNTADGEGAVDSNLNQNDEIPEPTQCSLMCPNECVMSWGLMSKCPSCDPHTMQ 1008
Db 648 VRCSSGRDED-----CAPAGREFCPARRCHL---RPCATWHSQWNSKSKRS----- 689
Qy 1009 RRTTHLLRPLSLNRTCAEDSQVOPCLLNENCFQOYNLTWSTCOLSENAPCGQGVTRTL 1068
Db 690 -----CGGSSVYRD 698
Qy 1069 LSCV-CSDGKPVSMOCBOHNLKPKQRM--SIPCLVECVNQLSGW--TAWTCSQTC 1122
Db 699 VQCVDRDLRLPLRPHFCQPGPAKPAHRPCGAQPCCL-----SWYTSSWRECEAC 748
Qy 1123 GHGGRMSRTEFIIMPTQGGRCPCPTLTQEKI--CPVTPCYSWVLGNWSACKLEGDCGE 1180
Db 749 GGG---EQORLVTCPEFG---LCEERALRENTTRPCNTHCTQWVVGPGWQC---SAPCG 799
Qy 1181 GVQIRSLSCWHSWGSISHAAGRVED--ALCGBMPQDSILKQLCSVPC-PGDCHLITE 1234
Db 800 GVQRRLLVKC-----VNTQTGLPEEDSDQCGHEAWPES-----SRPCGTEDCEPVE 844

```

Species: Caenorhabditis elegans  
Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
Accession: T25061  
Submitted to the EMBL Data Library, November 1995  
Reference number: Z19975  
Accession: T25061  
Status: preliminary; translated from GB/EMBL/DBJ  
Molecule type: DNA  
Residues: 1-788 <WIL>  
Cross-references: EMBL:Z68011; PIDN:CAA92014.1; GSPDB:GN00028; CESP:T21B6.3  
Experimental source: clone T21B6  
Genetics:  
Gene: CESP:T21B6.3  
Map position: X  
Introns: 20/1; 47/1; 76/1; 152/1; 735/2; 754/2

Query Match 4.1%; Score 340.5; DB 2; Length 788;  
Best Local Similarity 26.6%; Pred. No. 4.6e-14;  
Matches 120; Conservative 51; Mismatches 145; Indels 135; Gaps 25;

2y 773 PLVETELCPDEFISQYGNWSD---CIIPEGRRPHRLRVQADSKEGEGRLRFAVAC 829  
Db 441 PACDGGQCVNPPVVGWHDWSDWSTC-----SCTCGDGAQRREC 482

2y 830 SDKNGRPVDFSCSSSGYIQEKVI-PCPFDCKLSDWSSWGSSCGIGVIRSKWLKE 888  
Db 483 STNN-----COGADYETEPNGLPCO---TWSEWCWSTCSASCSSGQGRTRFC-- 529

2y 889 KPYNGRGPCKLDLKNQVHEAVPCYSECNQYVWVHNSCKINNELRLSLRGCGGTQSRK 948  
Db 530 --HLGNRCRKYSEQCSAGPC-FEWSQW---EDWQC-----SVTCGQGVAVRQ 575

2y 949 IRCVNTADGEGNAVDNLCNQDIPETOS-----CSLMCPNECWSEWGLWKCQPS 1001  
Db 576 RTCL-----GGVFGDLQO---GPKQRAQDGGPCLWSP-----WQENWSTCSAS 618

2y 1002 CDPHTMQRTRHLLRPLSLNRTCAEDSQV---QPCL-LNENCFOQYNLTEWSTQSLSEN 1057  
Db 619 CG-SGNKRQR-VQFGTDCQGNESQCYGPPCAETWEC-----EWSGC-----S 664

2y 1058 APGCGVTRLLSCVSDGKPVSMDCQBNLEKFORMSIPCLVECVNQCISGWTAMTE 1117  
Db 665 SKGPGQRTTRGLCPNGQEAAT-TCQGPSIE-----TTLCEGQSCCNWSEWCHWSM 715

2y 1118 CSOTCGHGGEMSTREI--IMPTQGEGRCPHETQETCPVTPCYSWVLGNWSACKLEG 1175  
Db 716 CDKECGG---QVRYEYFNFTGCEWSPCTOLA----- 746

2y 1176 GDCGEGVQIRSLSCMVHSGSISHAAGRVEDA 1206  
Db 747 --CEVGQSRROCVGESG--CHCIGLAES 773

RESULT 12  
T00260  
hypothetical protein KIAA0605 - human  
C:Species: Homo sapiens (man)  
C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 21-Jul-2000  
C:Accession: T00260  
C:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.  
DNA Res. 5, 31-39, 1998  
A:Title: Prediction of the coding sequences of unidentified human genes. IX. The complete  
A:Reference number: Z14086; MUID:98290545; PMID:9628581  
Accession: T00260  
Status: preliminary; translated from GB/EMBL/DBJ  
Molecule type: mRNA  
Residues: 1-951 <NAG>  
Cross-references: EMBL:AB011177; NID:G3043733; PIDN:BAA25531.1; PID:G3043734  
Experimental source: brain  
Genetics:  
Note: KIAA0605  
Superfamily: thrombospondin type 1 repeat homology

F:46-106/Domain: thrombospondin type 1 repeat homology &lt;THRL&gt;

Query Match 4.1%; Score 340.5; DB 2; Length 951;  
Best Local Similarity 18.4%; Pred. No. 5.6e-14;  
Matches 219; Conservative 117; Mismatches 357; Indels 499; Gaps 58;

Qy 200 QWSWSPCSKTCRSGSL-----PGRSRNRVKNMAIGGKSCP---EL 241  
Db 52 EWTKWTAFSRSCGGVTSQBRHCLQRRKSVPGNRTCTGTSKRYQLCRVQCPDGRS 111

Qy 242 LEKACTIVGELLQOCPRYSWRTESEKQCVSLLEQDDPHMHTVGVCGGIGTREVYC 301  
Db 112 PREQCYSFNSHV-----YNGRTHQWKL-----YDDYVHISKPC-----DLHC 152

Qy 302 AQ-----SVPAALRAKEVRPEKALCVGAPLPQLCNIPCTDCTIVSWSAWGL 354  
Db 153 TTVDGQRQLMVPARD-----GTSCKLTLDR--GV 179

Qy 355 CIHENCHEPCKKG--PRTRQRHVLMESTGPAHCPH-----LVESVPC- 396  
Db 180 CVSGKC-EPICDGVLFSTHTLQKCGICQDGGSCHTVTGNKGNHLYSLVTHIPAG 238

Qy 397 --EDPMCVRMLASEGICFPDGH-----KQGLG 421  
Db 239 ARDIQIVERKKSADVLALADEAGYFFENGYKVDSPKXFNIAGTVVKYRRPMDVIETGIE 298

Qy 422 HRLKAVC-----QNDRGEDVSGSLCPVPPPERKSCBPCPCMDCVLSEWESS 471  
Db 299 YIVAGGTNGLNVMWNQNGKSPITFYTLQPHESRPQI-----YIG 345

Qy 472 CSQCSNKNSDGK-----QTSRRTI--LALAGEGG--KPCPPQALQEH 512  
Db 346 FSESAGSLDGLAGLMGFPHNGSLYQASSERLGLDNLRFHGLDLMELGPQGETNE 405

Qy 513 LQND--HSCMLHWETSPWGPCSEDLTALNATIGWNGEAT----- 552  
Db 406 VCEQAGGAC-----EGPRGKGRDRNVTGTLTGDKDDEVDTHFASQEFFSANAISD 460

Qy 553 --CGVGIQTRVRFVKSHVGVMTKRCPDSTRPET-----VRPCFLPCKK 595  
Db 461 QLLGAGDLKD--FTLNVTNSIFAQAPRSLAESFVDYEENEGAGPYLLNSYLESS 519

Qy 596 DCIVTAFSEWTPCPR-----MCOAGNATVKQSYRIIIOBAANGOGCPDLYEERECED 650  
Db 520 DRVANSSE-APPNVSTSLTTSAGNRTHK--ARTR-----PKARKQG 559

Qy 651 VSLCPVYRWKPKWSPCLLVPSVWQGITGSSACGKGLQTRAVSCISDDNRAEWMECL 710  
Db 560 VSPADMYRWKLSSEPC-----SATCTTGVMSAYAMCVRYDGVEDV----- 600

Qy 711 KQTNGMPLLVQECTVPCREDCTFTAWSKFTPCSTNCEATKSRRLQTLGSRKKEKQDSD 770  
Db 601 -----DSYCD-----LTPRPVHEFCAGRE 621

Qy 771 LYPLVETELCPDEFISQYGNWSDCILPEGRPHRLRVQADSKEGEGRLRFAVACS 830  
Db 622 CQPRWET-----SSWSEC-----SRTCGEGYQFRVVR- 649

Qy 831 DKNGRPVDFSCSSSGYIQEKVIPCFFDCKLSDWSSWGSSCGIGVIRSKWLKEP 890  
Db 650 --WKMISPGF-DSVV-----SDLCEAAEAVRE----- 675

Qy 891 YNGRGPCKLDLKNQVHEAVPCYSECNQYVWVHNSCKINNELRLSLRGCGGTQSRKIR 950  
Db 676 -----ERKTCRNPACGPQWSEWSEC-----TAKGCE-----R 704

Qy 951 CVNTADGEGGAVDNLNQNDEIPEPETQCSLMCEWSEWGLWSKCPQSCDPHTMQR 1010  
Db 705 SVWTRDIR-----CSEDE-----XLCDDNT----- 724

Qy 1011 TRHLRLSLNRTCAEDSQVQPCLLNENCFOQYNLTEWSTQSLSENAPCGQGVTRLLS 1070  
Db 725 -----RP-VGEKNCTG-----PPC-----DRQWTVSDWGPC-----SGSCGQGRTRHVV 763

```

QY      852 C-VIPCPDFCKLSDWSSWGS CSSCGIGVIRIRSKWLKEKPYNGGRPCPKXLDLNQ-VHEA  909
Db      792 CNVATQNDGIWLSLWNSDCSRVCGKGLRSRSC-----FGSGWGASSEQFCNEQ  845
QY      910 VPCVSECNOY-SWVVEHWSSCKINNELRSIRFCGGTOSRKIRCVNTADGEGAVDSNLN  968
Db      846 ACASSANDWGTW--SGWSQC-----SVSCGAGVK-RRTRCTGNCPCGYKESALCN  895
QY      969 ODEIPPTQCSLMCPNE-CVMSEWGLWSKCPQSDPHMTQRRTRHLLRPLSLNSRTCAED  1027
Db      896 DRD-----CENKAAWGGWYSSCSETCGDGVK-----  926
QY      1028 SOVQCLINENCFOYNLTETWSTCOL 1054
Db      927 --VRKXSGNCDGOY---EKQVCNL 948

RESULT 14
T42215
zonadhesin - mouse
N:Alternate names: sperm-specific membrane protein
C:Species: Mus musculus (house mouse)
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T42215
R:GAO, Z.; Garbers, D.L.
J. Biol. Chem. 273, 3415-3421, 1998
A:Title: Species diversity in the structure of zonadhesin, a sperm-specific membrane pr
A:Reference number: Z22080; MUID:98123114; PMID:9452463
A:Accession: T42215
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-5376 <GAO>
A:Cross-references: EMBL:U97068; NID:G3327420; PID:G3337421; PIDN:AAC26680.1
C:Genetics:
A:Gene: Zan
A:Map position: 5
C:Function:
A:Description: functions in multiple cell adhesion processes
A:Note: found exclusively on the apical region of the sperm head
C:Keywords: cell adhesion

Query Match 4.1%; Score 335; DB 2; Length 5376;
Best Local Similarity 20.3%; Pred. No. 7.3e-13;
Matches 325; Conservative 154; Mismatches 523; Indels 596; Gaps 97;

QY      18 CHHFALQPTTEQ---ACLIIPRCDCVVEFLPWNCSKGCGKLOHRTRAVIAPLFGGL 74
Db      2933 CATIITLQCPAHSHTNCLPPCQPSCLDSE----GHC----- 2964
QY      75 QCPNLTESRACDAPISCLPG--BEEY-TFSLKVGPMKCLRPLHKEI-NPSGRVTLDNF 129
Db      2965 -----EGSTYKASACQCGVCVPDYVLNNKCVPRIECGCKDAQGVLPADKTIW--N 3016
QY      130 SDSNERVTFKH---QSYKAHHHSKSWAIEIGYQTRQVSCSTRSDGQ---NAMLSLCLQDSF 183
Db      3017 RGCTQSTCTCKGAIQCQFQCFSEYTKXDI--EDGNSNCTEISLQCPANSNFTSLCPSCQ 3074
QY      184 PLTVQSCI-MPKDCBTSQSSWSWSCSK--TCRSGLLPGRFSRNRNVKHMAGGKCEPE 240
Db      3075 P-----SCNTDVHCBSGSPNTLSRCRECVCSG-----Y 3105
QY      241 LLEKACIVBEGELQQCPRYSWRTSEWKEQVSLLEQQDPHHNHTGPVC-----GGG 293
Db      3106 VLHNDKCI-----LRNQ-----GCKDAQGALIPGKLT-WITSG--CTQSCNCTGA 3149
QY      294 ICTR-----EYVC-----AQSVFAAAALRAKEYS-----RPV---EKALCVGPAP 330
Db      3150 IQCNFQCPKTYCKDLKDGSSNCTNIPLQCPAHSRYTNCLPSPCPICLDEGLCEGTSP 3209
QY      331 -LPSQLCNIPGSTDICVSSWSAWGLCIHEN-----CHEPQKKGFRR-----QRH 375
Db      3210 KVPS-----TCREGICOP-----GYLMHKNKCVLRIFGCKNTQG--AFISADKTWISRG 3258

```





```

QY 979 CSLMCPNECVMSKCPQSDPH--TWQRTRHLLRPLSLNRITCAEDSQVQPCLLN 1036
Db 1001 C---CPETGI---WSDWTSTENQCRDYCGSGNQTR-----TRICTSDADGCPQ--- 1043
QY 1037 ENCQFQYNLTENSTCQLSENAPCGQGVTRLLSCV-----CSDGKPVSMDOCEOH----- 1087
Db 1044 ----CGPTTITE-----PCGTGV-----CYFRLSCCPGYTATVEG-NQHICGP 1082
QY 1088 ---NLEKFPQMSIPCLVECVVNCQLSG-WTAWTECSQTCGHGGRMSRTRFIIMPTQG--- 1140
Db 1083 LTTAVADPDKLN-TCGVSCCPFSAGI-WGEWVSVSGNDTCGSCQGETRKRKCLSLQYGCAC 1141
QY 1141 EGRCPTELTQETC-----PVTPCYS----- 1162
Db 1142 TGNATDTSVCASSVCLPRTSCCTGFKOMNITGRIFYCGPLPVVPAFNPETTCDDPEK 1201
QY 1163 -----WVLGNWSACKLEGDCGEGVQIRSLSCMVHSGSISHAAGRVEDALCGEMPFODS 1216
Db 1202 TGLMNDW--GAWTTCSATCGGC--GTQTRSKT-----CASAPY--- 1235
QY 1217 ILKQLCSVPCPDCHLTWSEWSSTCELTCDIGRSFETVGRQSRRTFIIOSFENQDSCPQ 1276
Db 1236 -----GCPCTGD--LTETQ--SCAKQVC-----TTGAQCCAGKEVATGYDGAQYCOD 1278
QY 1277 QVLETRPCTGCKYHYTWKASLNNNNERTVWCQRSDGV--NVTGCGSPQARPAIROCIP 1334
Db 1279 NTPEV--CTG-----TW-----TEWATLEGAVCNDTCGNCG-----IIP 1310
QY 1335 ACRKPFSTYCTGGVCGCEKGYT 1356
Db 1311 TSR-----YCFPSG-CQCSGAVT 1327

```

Search completed: February 24, 2004, 01:08:25  
Job time : 67.0517 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

MM protein - protein search, using sw model

Run on: February 23, 2004, 20:35:55 ; Search time 41.0236 Seconds  
(without alignments)  
1859.483 Million cell updates/sec

Title: US-10-022-710-2

Effect score: 8241

Sequence: 1 MVRCIQLNRTVVAIVEICEH.....QSTPPQKPLTLAYDGLDM 1465.

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	664	8.1	1935	1	AT99 HUMAN
2	628.5	7.6	1906	1	AT20 MOUSE
3	632	7.5	1911	1	AT20 HUMAN
4	428.5	5.2	807	1	FSP0 RAT
5	393.5	4.8	803	1	FSP0 XENLA
6	360	4.4	1593	1	AT12 HUMAN
7	358.5	4.4	1093	1	SM5B MOUSE
8	357.5	4.3	1700	1	BAR3 CHITE
9	351.5	4.3	1077	1	SMSA HUMAN
10	345.5	4.2	1074	1	SMSA HUMAN
11	336.5	4.1	1093	1	SM5B HUMAN
12	335	4.1	5376	1	ZAN MOUSE
13	334	3.9	867	1	SSPO BOVIN
14	323	3.9	1095	1	AT17 HUMAN
15	318.5	3.9	1077	1	AT10 HUMAN
16	314.5	3.8	1210	1	AT19 MOUSE
17	312.5	3.8	469	1	PROF MOUSE
18	310.5	3.8	2321	1	NTC3 HUMAN
19	309	3.7	2318	1	NTC3 MOUSE
20	306	3.7	1224	1	AT16 HUMAN
21	305	3.7	2871	1	FBN1 BOVIN
22	304.5	3.7	1207	1	AT19 HUMAN
23	303	3.7	2703	1	NOTC DRONE
24	302.5	3.7	1584	1	BAIL HUMAN
25	301	3.7	2871	1	FBN1 PIG
26	300.5	3.6	1696	1	PK55 BRACL
27	298	3.6	2437	1	NTC1 BRARE
28	297.5	3.6	2319	1	NTC3 RAT
29	297.5	3.6	2911	1	FBN2 HUMAN
30	297	3.6	2907	1	FBN2 MOUSE
31	292	3.5	2531	1	NTC1 MOUSE
32	288.5	3.5	1223	1	AT14 HUMAN
33	285.5	3.5	470	1	PROF_CAVPO

## RESULT 1

ID	AT99_HUMAN	STANDARD	PRT	1935 AA
AC	Q9P2N4; Q9NR29;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	10-OCT-2003 (Rel. 42, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	ADAMTS-9 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 9) (ADAM-TS 9) (ADAM-TS9).			
GN	ADAMTS9 OR KIAA1312.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
[1]	SEQUENCE FROM N.A. (ISOFORM 3).			
RP	TISSUE=Fetal;			
RC	MEDLINE=20396138; PubMed=10936055;			
RX	Clark M.E., Kelnner G.S., Turbeville L.A., Boyer A., Arden K.A., Maki R.A.;			
RA	"ADAMTS 9, a novel member of the ADAM-TS/Metallospodin gene family.";			
RT	Genomics 67:343-350(2000).			
RL	[2]			
RN	SEQUENCE FROM N.A. (ISOFORM 1), AND FUNCTION.			
RP	MEDLINE=22513925; PubMed=12514189;			
RX	Somerville R.P., Longpre J.-M., Jungers K.A., Engle J.M., Ross M., Evanko S., Wight T.N., Leduc R., Apte S.S.;			
RA	"Characterization of ADAMTS-9 and ADAMTS-20 as a distinct ADAMTS subfamily related to Caenorhabditis elegans GON-1.";			
RT	J. Biol. Chem. 278:9503-9513(2003).			
RL	[3]			
RN	SEQUENCE OF 159-1935 FROM N.A. (ISOFORM 2).			
RP	TISSUE=Brain;			
RC	MEDLINE=20181126; PubMed=10718198;			
RX	Nagase T., Kikuno R., Ishikawa K.-I., Hirokawa M., Ohara O.;			
RA	"Prediction of the coding sequences of unidentified human genes. XVI. The complete sequences of 150 new cDNA clones from brain which code for large proteins in vitro.";			
RT	DNA Res. 7:65-73(2000).			
RL	CC -!- FUNCTION: Cleaves the large aggregating proteoglycans, aggrecan and versican.			
CC	CC -!- CATALYTIC ACTIVITY: Cleaves aggrecan at the 1838-Glu- -Ala-1839 site and versican at the 1428-Glu- -Ala-1429 site.			
CC	CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).			
CC	CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (By similarity).			
CC	CC -!- ALTERNATIVE PRODUCTS:			
CC	CC Event=Alternative splicing; Named isoforms=3;			
CC	CC Name=1; Synonyms=ADAMTS-9B;			
CC	CC IsoId=Q9P2N4-3; Sequence=Displayed;			
CC	CC Name=2; Synonyms=long;			
CC	CC IsoId=Q9P2N4-1; Sequence=VSP_007548; VSP_007549;			
CC	CC Note=May result from the retention of an intron in the cDNA leading to a premature stop codon;			
CC	CC Name=3; Synonyms=Short;			

P58459 mus musculus  
Q07008 rattus norv  
P46531 homo sapien  
P35555 homo sapien  
O15072 homo sapien  
P11880 mus musculus  
Q04592 mus musculus  
Q8C919 mus musculus  
P07996 homo sapien  
Q61554 mus musculus  
P35448 xenopus lae  
P21783 xenopus lae

## ALIGNMENTS

Isoid-Q9P2N4-2; Sequence=VSP\_005499, VSP\_005500;  
-!- TISSUE SPECIFICITY: Highly expressed in all fetal tissues.  
Expressed in a number of adult tissues with highest expression in  
heart, placenta and skeletal muscle.  
-!- DOMAIN: The spacer domain and the TSP type-1 domains are important  
for a tight interaction with the extracellular matrix (by  
similarity).  
-!- DOMAIN: The ancillary domains, including the TSRs domain, are  
required for specific extracellular localization and for its  
versicanase and aggrecanase activities.  
-!- PM: The precursor is cleaved by a furin endopeptidase (by  
similarity).  
-!- SIMILARITY: Belongs to peptidase family M12B.  
-!- SIMILARITY: Contains 1 disintegrin-like domain.  
-!- SIMILARITY: Contains 1 GON domain.  
-!- SIMILARITY: Contains 15 TSP type-1 domains.  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; AF261918; AAF89106.1; -  
EMBL; AF488803; AAO5765.1; -  
DR EMBL; AB037733; BAA92550.1; -  
DR HSSP; P15167; IATL.  
DR MEROPS; M12.021; -  
DR Genew; HGNC:13202; ADAMTS9.  
MIM; 605421; -  
DR GO; GO:0008217; F:metalloproteinase activity; TAS.  
DR GO; GO:0007275; P:development; TAS.  
DR GO; GO:0006516; P:glycoprotein catabolism; TAS.  
DR InterPro; IPR001762; Disintegrin.  
DR InterPro; IPR001818; Pept\_M10A\_M12B.  
DR InterPro; IPR006025; Pept\_M\_Zn\_BS.  
DR InterPro; IPR001590; Peptidase\_M12B.  
DR InterPro; IPR002870; Peptidase\_M12B\_N.  
DR InterPro; IPR000884; TSP1.  
DR Pfam; PF01562; Pcp\_M12B\_propep; 1.  
DR Pfam; PF01421; Reprolysin; 1.  
DR Pfam; PF00090; TSP1; 11.  
DR SMART; SMO0209; TSP1; 12.  
DR PROSITE; PS00215; ADAM\_MEROPS; 1.  
DR PROSITE; PS00546; CYSTEINE SWITCH; FALSE NEG.  
DR PROSITE; PS00427; DISINTEGRIN\_1; FALSE NEG.  
DR PROSITE; PS00214; DISINTEGRIN\_2; FALSE NEG.  
DR PROSITE; PS00092; TSP1; 14.  
DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;  
Repeat; Extracellular matrix; Alternative splicing.  
FT SIGNAL 1 18 POTENTIAL.  
FT PROPEP 19 287  
FT CHAIN 288 1935  
FT DOMAIN 293 499 ADAMTS-9.  
FT DOMAIN 509 587 METALLOPROTEASE.  
FT DOMAIN 588 643 DISINTEGRIN-LIKE.  
FT DOMAIN 644 752 TSP TYPE-1 1.  
FT DOMAIN 753 877 CYS-RICH.  
FT DOMAIN 878 936 SPACER.  
FT DOMAIN 939 997 TSP TYPE-1 2.  
FT DOMAIN 998 1049 TSP TYPE-1 3.  
FT DOMAIN 1052 1109 TSP TYPE-1 4.  
FT DOMAIN 1110 1166 TSP TYPE-1 5.  
FT DOMAIN 1182 1240 TSP TYPE-1 6.  
FT DOMAIN 1241 1296 TSP TYPE-1 7.  
FT DOMAIN 1328 1379 TSP TYPE-1 8.  
FT DOMAIN 1382 1440 TSP TYPE-1 9.  
FT DOMAIN 1441 1494 TSP TYPE-1 10.  
FT DOMAIN 1497 1555 TSP TYPE-1 11.  
FT DOMAIN 1556 1611 TSP TYPE-1 12.  
FT DOMAIN 1557 1611 TSP TYPE-1 13.

FT	DOMAIN	1612	1676	TSP TYPE-1 14.
FT	DOMAIN	1677	1734	TSP TYPE-1 15.
FT	DOMAIN	1735	1935	GON.
FT	DOMAIN	88	96	POLY-SER.
FT	SITE	223	223	CYS-STEINE SWITCH (POTENTIAL).
FT	METAL	434	434	ZINC (CATALYTIC) (BY SIMILARITY).
FT	ACT SITE	435	435	BY SIMILARITY.
FT	METAL	438	438	ZINC (CATALYTIC) (BY SIMILARITY).
FT	METAL	444	444	ZINC (CATALYTIC) (BY SIMILARITY).
FT	CARBOHYD	112	112	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	135	135	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	271	271	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	749	749	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	840	840	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1213	1213	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1267	1267	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1788	1788	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1806	1806	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	VARSPLIC	1064	1072	CLVTCGKCH -> VRWEGCYFP (in isoform 3).
FT	VARSPLIC	1073	1935	/FTid=VSP_005499.
FT	VARSPLIC	1624	1629	Missing (in isoform 3).
FT	VARSPLIC	1630	1935	/FTid=VSP_005500.
FT	VARSPLIC	1630	1935	CSVTG -> VPSWEL (in isoform 2).
FT	VARSPLIC	1630	1935	Missing (in isoform 2).
FT	CONFLICT	46	46	S -> G (IN REF. 1).
FT	CONFLICT	96	96	P -> S (IN REF. 1).
FT	CONFLICT	182	182	D -> G (IN REF. 2).
FT	CONFLICT	367	367	F -> L (IN REF. 1).
FT	CONFLICT	1117	1117	V -> G (IN REF. 3).
SQ	SEQUENCE	1935	AA; 216556	MM; FD3D51E88300A3C6 CRC64;

Query Match 8.1%; Score 664; DB 1; Length 1935;  
Best Local Similarity 21.3%; Fred. No. 1.7e-36;  
Matches 289; Conservative 123; Mismatches 42; Indels 526; Gaps 70;

QY	201	WSS	---WSPCKTCTSGSLLPGRSRNRNKHMTAIGGKCEKPELLEKACIVEGELL---	254
DB	881	WNSHGPWQACSKPCQ	-----GERKR-----	914
QY	255	-QCPRYSWRSEWKEQVSLLEQDPHVV	-----GPVGGGIQTRVYCAQSVAA	308
DB	915	DQCDRLPQPGHITPCGTDC	-----DLRHWVNRSECSAQCGLYRTLDIYCAK	964
QY	309	AALRAKEVSRPVEKALCVGPAPLPOLCNI	PCSTDCIVSSW--SANGLCIHENCHPOCK	366
DB	965	-YSRLDGKTEKVDGFC	-SSHPKPSN--REKSGECNTGGRYSATWC	1009
QY	367	KGFRTRQRHVLWESTGPAGHCPHLVESVPCED	PMCYRWLASEGICFPDGHGKGLGHLRLK	426
DB	1010	-----	-----SKSCDGTQRR	1021
QY	427	AVCONDRGEDVSGSLCPVPPPPKSC	-EIPCRMDCVLSEWT--EWSSCSOSCSNKNSDG	483
DB	1022	AICVTRNDVLDSSKTHQEKVTIQRCEFFC	-----PQKSGDWSCLVTC	1069
QY	484	KQTRERTILALAGE	---GGKPCPPSQALQEHRLNDSHSCMLHMETSPWGPCSEDTLWTA	540
DB	1070	KGHKEQVWCQFGEDRLNDRMCDPETHKPTK	SMQTCQPECAS--WOAGPWQCS	1120
QY	541	LNATIGWNGEATCGVGIQTRVFC	-VKSHVGQVMTKRCFDSRTPETVPCFLPCKDC	597
DB	1121	-----	-----VTGQGYQLRAVKCIITGYNVDDNDCNAATRTDQDCELP	1167
QY	598	-----	-----IVTAFSEMTPCPRVQAGNATVKQSRYLIIQEAAGQCEPDT	641
DB	1168	PPAAPETRETSYAPRTQWRFGSWTFCATCGK	-----TMRVYVSCRDENG	1215
QY	642	LYEBREC	-----EDVSLCPVTRWPKQKWSPCILVPSVWQIGTSSACGKGLQTR	692
DB	1216	VADESACATLPRFAKEKSCVTPCGQKALDWSSC	-----SVTCQGRATR	1261

```
693 AVSCTSDNRSAMMECKLQTMGMPLLVQECTVPCREDCTFTAMSKFTPCSTNCEATKSR 752
b 1262 QWCVNYSVHVIDRSEC--DDYIETDQCSM-----SFC----- 1295
y 753 RRLQTKSRKKEKCOOSDL--YPLVETELCPDEFISOPY-----GNWSDCILPEG 801
b 1296 -----PORTPDSGLAQHPQFQNEYRPRASPSRTHVLGNQWRTGPGAC----- 1340
y 802 RREPHRGLRVQADSKCEGSGLEFRFVACSDKNG-----RPVDFSCSSSSYIQEK 851
b 1341 -----SSTCAGSRRVVVVCODENGYTANDCVERIKDEQRAESG----- 1381
y 852 CVIPCFDCKLSDWS--SWGSSSSCGIGVIRSRKWLKEKPYNGRCPKLDLKNQVHEA 909
b 1382 ---PCP---QWAGNNGEETKLCGGGIRTR---LVVQQRNGERFP--DLSCEILDK 1427
y 910 VPCYSECNOY-----SWVEHWSCKINNELSRLCGGGTQSRKIRCVNTADGGGAVD 963
b 1428 PPDREOCNTHACPHDAWSTGPWSSC-----SVSCGRGHQORVYCMAX---DGSLE 1477
y 964 SNLCNQDEIPEPTQSC--SLMCPNECVMSW--GLMSKCPQSCDPTMQRTRHLLRPSLN 1020
b 1478 SDYCKHLAKPHGRKCRGERCP-----KWKAGAWSQCSVSCGRGVQQRH---VGCCIG 1527
y 1021 SRTCAEDSQVQCLLNEN-----CFQFYNLTEWSTCQLSENAPCGQGVTRFLLSCV 1072
b 1528 THKIARETECNPTYPRESRDCQGRCPPLYTWRABEWOEC-----TKTCGEGSRVKKVCV 1583
y 1073 -----CSDGK--PVSMDQCEQHNLEKQFQMSIPCLVCEVNVNQLSGWTA--WTEC 1118
b 1584 DDNQNBVHGARDVSRPVDRESLSIQ-----PC--EYV-----WITGEWSEC 1624
y 1119 SQTCHGGRMSRTRFIIMTQEGRCPTELTOEKT--CP-----VTPCY-----S 1162
b 1625 SVTCGKGYQRLVSCSEIYTGKEN-----YEYGYQTINCPGTQPPSVHPCYLRECPVSAT 1680
y 1163 WVLGNWSACKLEGGDCGEGVQVRLSRCSMVHSGSISHAA----- 1200
b 1681 WRVGNWGSV---SCGVGMQSRVQCLTNEQPHLCHTDLKPERKTCRVNVCNLPQ 1737
y 1201 -----GRVEDALCGE--MPFQDSILKQLCVPCPGDCHLTWSEWSTCELCIDGR 1249
b 1738 NCKEVKRLKAGSED---GEYFLMIRGKLLKIFCA-----GMESHPKXVTV--LVHGDSE 1787
y 1250 SP-ETVGRQSRSTFIIOFENQDSC-----PQOVLEI----- 1281
b 1788 NFSEVYGHRLHNPTECPYNGSRDDCCQKQVTAAGFSFKIRIDLTSMQIITDLOFA 1847
y 1282 ----RP---CTGGKCYHYT-----WKASLWNN-----NERTVWCQRS-----D 1312
b 1848 RTSEGHVPVPFATAGDCYSAKCPQGRFSLNLYGTGLSLTESARWISQNGYAVSDIKKSPD 1907
y 1313 GUNVTGCGSPQAPPAIROCIACRKPFSYCTCGGVCG 1350
b 1908 GTRVWGC-----GGYCG 1920
```

## RESULT 2

```
AT20_MOUSE
ID AT20_MOUSE STANDARD; PRT; 1906 AA.
AC P59511;
JT 10-OCT-2003 (Rel. 42, Created)
JT 10-OCT-2003 (Rel. 42, Last sequence update)
JT 15-MAR-2004 (Rel. 43, Last annotation update)
DE ADAMTS-20 precursor (EC 3.4.24.-) (A disintegrin and
DE metalloproteinase with thrombospondin motifs 20) (ADAM-
DE TS20).
EN ADAMTS20.
EN Mus musculus (Mouse).
XC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
XC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
XC NCBI_TaxID=10090;
[1]
```

```
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Fetal brain;
RX MEDLINE=22566039; PubMed=12562771;
RA Llamazares M., Cal S., Quesada V., Lopez-Otin C.;
RT "Identification and characterization of ADAMTS-20 defines a novel
RT subfamily of metalloproteinases-disintegrins with multiple
RT thrombospondin-1 repeats and a unique GON domain.";
RL J. Biol. Chem. 278:13382-13389(2003).
[2]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND DISEASE.
RC STRAIN=DEA/2;
RX MEDLINE=22806432; PubMed=12925592;
RA Rao C., Foerzler D., Loftus S.K., Liu S., McPherson J.D.,
RA Jungers K.A., Apte S.S., Pavan W.J., Beier D.R.;
RT "A defect in a novel ADAMTS family member is the cause of the belted
RT white-spotting mutation.";
RL Development 130:4665-4672(2003).
CC -!- FUNCTION: May play a role in tissue-remodeling process occurring
CC in both normal and pathological conditions.
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1; Synonyms=ADAMTS20 B long isoform;
CC IsoId=P59511-1; Sequence=Displayed;
CC Name=2; Synonyms=ADAMTS20 A short isoform;
CC IsoId=P59511-2; Sequence=VSP_007606, VSP_007607;
CC -!- TISSUE SPECIFICITY: Expressed at low level in testis and brain.
CC -!- PTM: The precursor is cleaved by a furin endopeptidase (By
CC similarity).
CC -!- DISEASE: Defects in ADAMTS20 are the cause of the belted (bt)
CC phenotype. It is a pigmental defect which occurs as a result of a
CC defect in melanocyte development.
CC -!- SIMILARITY: Belongs to peptidase family M12B.
CC -!- SIMILARITY: Contains 1 disintegrin-like domain.
CC -!- SIMILARITY: Contains 1 GON domain.
CC -!- SIMILARITY: Contains 15 TSP type-1 domains.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ512753; CAD54808.3; -
CC EMBL; AY189815; AA074895.1; -
CC EMBL; AY189816; AA074896.1; -
CC MGI; MGI:2660628; Adamts20.
CC InterPro; IPR001762; Disintegrin.
CC InterPro; IPR006025; Pept_Mn_Bs.
CC InterPro; IPR001590; Peptidase_M12B.
CC InterPro; IPR002870; Peptidase_M12B_N.
CC InterPro; IPR000884; TSP1.
CC Pfam; PF01562; Pep_M12B_propep; 1.
CC Pfam; PF01421; Reptolysin; 1.
CC Pfam; PF00050; tsp_1; 11.
CC SMART; SM00209; TSP1; 14.
CC PROSITE; PS02015; ADAM_MEPRO; 1.
CC PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
CC PROSITE; PS0214; DISINTEGRIN_2; FALSE_NEG.
CC PROSITE; PS0092; TSP1; 13.
CC PROSITE; PS00143; ZINC_PROTEASE; 1.
CC Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
CC Repeat; Extracellular matrix; Alternative splicing.
FT SIGNAL 1 26
FT PROPEP 27 249
FT CHAIN 250 1906
FT DOMAIN 250 464
FT DOMAIN 465 552
FT DOMAIN 553 608
FT TSP TYPE-1 1.
```

FT	DOMAIN	609	720	CYS-RICH.
FT	DOMAIN	721	842	SPACER.
FT	DOMAIN	843	901	TSP TYPE-1 2.
FT	DOMAIN	902	962	TSP TYPE-1 3.
FT	DOMAIN	962	1015	TSP TYPE-1 4.
FT	DOMAIN	1017	1074	TSP TYPE-1 5.
FT	DOMAIN	1075	1131	TSP TYPE-1 6.
FT	DOMAIN	1148	1202	TSP TYPE-1 7.
FT	DOMAIN	1203	1260	TSP TYPE-1 8.
FT	DOMAIN	1300	1351	TSP TYPE-1 9.
FT	DOMAIN	1354	1411	TSP TYPE-1 10.
FT	DOMAIN	1412	1465	TSP TYPE-1 11.
FT	DOMAIN	1468	1526	TSP TYPE-1 12.
FT	DOMAIN	1527	1584	TSP TYPE-1 13.
FT	DOMAIN	1585	1648	TSP TYPE-1 14.
FT	DOMAIN	1650	1706	TSP TYPE-1 15.
FT	DOMAIN	1707	1906	GON.
FT	METAL	399	399	ZINC (CATALYTIC) (BY SIMILARITY).
FT	ACT SITE	400	400	BY SIMILARITY.
FT	METAL	403	403	ZINC (CATALYTIC) (BY SIMILARITY).
FT	METAL	409	409	ZINC (CATALYTIC) (BY SIMILARITY).
FT	CARBOHYD	52	92	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	221	221	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	714	714	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	738	798	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	805	805	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1057	1057	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1562	1562	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1719	1719	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1759	1759	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1777	1777	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	VARSPLIC	1424	1425	CS -> VR (in isoform 2). /Ftid=VSP_007606.
FT	VARSPLIC	1426	1906	Missing (in isoform 2). /Ftid=VSP_007607.
FT	CONFLICT	1211	1211	D -> Y (IN REF. 2).
FT	CONFLICT	1262	1262	S -> L (IN REF. 2).
FT	SEQUENCE	1906 AA;	212041 MW;	0EB2568547B557EA CRC64;
SQL				
Query Match				
Best Local Similarity 7.6%; Score 628.5; DB 1; Length 1906;				
Matches 377; Conservative 163; Mismatches 558; Indels 691; Gaps 96;				
QY	85	CDAPTSCPLGSEE-----YTFSLKVG-----	105	
Db	377	CDPSRCSISEENGLSAAFTIAHELGHVFVNPHDDSPFKCEAGIKHQHVMAPTLVNHTS	436	
QY	106	--PWSKRLPLHKEI-----NPSRTV-----LDNSDSNE	134	
Db	437	PWTWACSOKHITEFLDTGHGECLLPNGRTVDLSPLPGSYVDGNROCELMFGFSQV	496	
QY	135	RVTFXHQSHKAHHHSKSWAIEIGYQTRQVSCTRSQGONAMLSLCLQDSPLT--VOSC---	190	
Db	497	CPYLKH-----CRLMCTSAGVH---KGRCTQHMLADGTSGCGP	534	
QY	191	-----TMPDCET-----SQSSWSNP---CSKTCSGSLLPG---FRSRNRNVKHAIG	233	
Db	535	XHCHRGLCVTRDMETRPVDGWGPWGPGYSCSRCTGGGKSTARLCDREPEENGRCYCVG	594	
QY	234	-----GGKECP-----ELLEKEACTIVEG-----LLQCPCRYSW-----	262	
Db	595	RRMKFRSNTDSCPKGKEDPREKQCSDFDKGFINDGLPNFWLFPKYSGIAVKDRCKLY	654	
QY	263	-----RTSEWKSCQSVLL-----LEQODPHWHVTP-----VCGG-----	292	
Db	655	CRVAGTTFYQLKDRVADGTSFGTETNDICVOGLCRQAQGDCHVLNSKAKRDCRGVC	714	
QY	293	-----GIQTREVY-----CAQSVPAAA-----	309	
Db	715	SSCQTLAGVFNSAHYGVVVVKIPAGATNIELLOHSYSGRPEDDNVYALSDTDQNFLNG	774	
QY	310	-----ALRAKEV-----SRPEKALCVGPAPLPS-----QLCN	337	

```
1295 KASLMNNERTWICOR-----SDG-----VNVGTG-----GCSPQ--ARPAAIRQCIPAC 1336
1703 -----NDCKLLATCKELQVNTVNRKGDYDLNVRGRLKIHCSGMQLENP---REYLPVLV 1754
1337 RKPPSY-----CTGGV-----CGCEK-----GYTEINKSGFLDYCKVPGS 1374
1755 KEDNPFSEIYGLRQNPVECPFGRRPRDCACENDYLPAGYTVFSKVR-----VDLESM 1808
1375 EDKQADV-----KNLSGKNRP-----VNSKIHDFKGSLO-- 1405
1809 QIKTADLLFSQTLGKGVFFATAGDCYSAARCPQGFPSINLAGTCMKISNTAK-WLAQGR 1867
1406 -----PLDPGRVKIYGVGGGFLIMIFLFTSYLVCKPKPKHQST 1448
1868 YASVVIHRSQDGTK---VYGRCGG-----FCGKCIPIHMAT 1899

RESULT 3
T20 HUMAN
C AT20 HUMAN STANDARD; PRT; 1911 AA.
C P59510;
T 10-OCT-2003 (Rel. 42, Created)
T 10-OCT-2003 (Rel. 42, Last sequence update)
T 10-OCT-2003 (Rel. 42, Last annotation update)
T ADAMTS-20 precursor (EC 3.4.24.-) (A disintegrin and
T metalloproteinase with thrombospondin motifs 20) (ADAM-TS 20) (ADAM-
T TS20).
E
N ADAMTS20.
S Homo sapiens (Human).
X Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
X Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
X NCBI_TaxID=9606;
X [1]
N SEQUENCE FROM N.A. (ISOFORM 1).
X MEDLINE=22513925; PubMed=12514189;
A Somerville R.P., Longpre J.-M., Jungers K.A., Engle J.M., Ross M.,
A Evanko S., Wright T.N., Leduc R., Apte S.S.;
T "Characterization of ADAMTS-9 and ADAMTS-20 as a distinct ADAMTS
T subfamily related to Canorhabditis elegans GON-1.";
L J. Biol. Chem. 278:9503-9513 (2003).
N [2]
N SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
C TISSUE=Liver;
X MEDLINE=22566039; PubMed=12562771;
X Llamazares M., Cal S., Quesada V., Lopez-Otin C.;
T "Identification and characterization of ADAMTS-20 defines a novel
T subfamily of metalloproteinases-disintegrins with multiple
T thrombospondin-1 repeats and a unique GON domain.";
L J. Biol. Chem. 278:13382-13389 (2003).
C -!- FUNCTION: May play a role in tissue-remodeling process occurring
C in both normal and pathological conditions.
C -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
C -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
C matrix (By similarity).
C -!- ALTERNATIVE PRODUCTS:
C Event=Alternative splicing; Named isoforms=2;
C Name=1;
C ISOID=P59510-1; Sequence=Displayed;
C Name=2;
C ISOID=P59510-2; Sequence=VSP_007106, VSP_007107, VSP_007108;
C -!- TISSUE SPECIFICITY: Very sparingly expressed, although is detected
C at low levels in testis, prostate, ovary, heart, placenta, lung
C and pancreas. Overexpressed in several brain, colon and breast
C carcinomas.
C -!- PTM: The precursor is cleaved by a furin endopeptidase (By
C similarity).
C -!- SIMILARITY: Belongs to peptidase family M12B.
C -!- SIMILARITY: Contains 1 disintegrin-like domain.
C -!- SIMILARITY: Contains 1 GON domain.
C -!- SIMILARITY: Contains 15 TSP type-1 domains.
C -!- SIMILARITY: Contains 15 TSP type-1 domains.
C This SWISS-PROT entry is copyright. It is produced through a collaboration
C -----
```

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF488804; AAC15766.1; -.
CC EMBL; AJ515153; CAD56159.3; -.
CC EMBL; AJ515154; CAD56160.2; -.
CC Genew; HGNC:17178; ADAMTS20.
CC InterPro; IPR001762; Disintegrin.
CC InterPro; IPR006025; Pept_M_2n_BS.
CC InterPro; IPR001590; Peptidase_M12B.
CC InterPro; IPR002870; Peptidase_M12B_N.
CC InterPro; IPR000884; TSPI.
CC Pfam; PF01562; Pep_M12B_propep; 1.
CC Pfam; PF01421; Reptolysin; 1.
CC Pfam; PF00030; TSP1; 11.
CC SMART; SMO0209; TSP1; 12.
CC PROSITE; PS0215; ADAM_MEPRO; 1.
CC PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
CC PROSITE; PS0214; DISINTEGRIN_2; FALSE_NEG.
CC PROSITE; PS0092; TSPI; 12.
CC PROSITE; PS00142; ZINC_PROTEASE; 1.
CC Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
CC Repeat; Extracellular matrix; Alternative splicing.
CC SIGNAL 1 21
CC PROPEP 22 253
CC CHAIN 254 1911
CC DOMAIN 254 467
CC DOMAIN 468 555
CC DOMAIN 556 611
CC DOMAIN 612 723
CC DOMAIN 724 846
CC DOMAIN 847 905
CC DOMAIN 906 962
CC DOMAIN 967 1024
CC DOMAIN 1025 1074
CC DOMAIN 1077 1136
CC DOMAIN 1153 1207
CC DOMAIN 1208 1265
CC DOMAIN 1305 1357
CC DOMAIN 1359 1417
CC DOMAIN 1418 1476
CC DOMAIN 1477 1532
CC DOMAIN 1536 1589
CC DOMAIN 1590 1653
CC DOMAIN 1655 1711
CC DOMAIN 1712 1911
CC METAL 403 404
CC ACT SITE 404 404
CC METAL 407 407
CC METAL 413 413
CC CARBOHYD 92 92
CC CARBOHYD 191 191
CC CARBOHYD 445 445
CC CARBOHYD 702 702
CC CARBOHYD 717 717
CC CARBOHYD 728 728
CC CARBOHYD 809 809
CC CARBOHYD 870 870
CC CARBOHYD 1062 1062
CC CARBOHYD 1457 1457
CC CARBOHYD 1543 1543
CC CARBOHYD 1573 1573
CC CARBOHYD 1764 1764
CC CARBOHYD 1782 1782
CC CARBOHYD 1853 1853
CC VARSPLIC 683 693
CC -----
CC /FTID=VSP_007106.
CC CSASCGKGRKYREVFCDQFQKLEDTNCSQVQKPPTHKAC
CC -----
```





```
CC cord and the PNS.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: Expressed at high levels in the floor plate.
CC -!- SIMILARITY: Contains 6 TSP type-1 domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M88469; AAA41174.1; -.
CC PIR; A38152; A38152.
CC InterPro; IPR002861; Reeler.
CC InterPro; IPR000884; TSP1.
CC Pfam; PF02014; Reeler; 1.
CC Pfam; PF00090; TSP1; 6.
CC SMART; SM00209; TSP1; 6.
CC PROSITE; PS50092; TSP1; 6.
CC Glycoprotein; Signal; Repeat; Cell adhesion.
CC SIGNAL 1 28 POTENTIAL.
CC CHAIN 29 807 F-SPONDIN.
CC DOMAIN 442 495 TSP TYPE-1 1.
CC DOMAIN 501 555 TSP TYPE-1 2.
CC DOMAIN 558 611 TSP TYPE-1 3.
CC DOMAIN 614 666 TSP TYPE-1 4.
CC DOMAIN 668 721 TSP TYPE-1 5.
CC DOMAIN 754 806 TSP TYPE-1 6.
CC CARBOHYD 214 214 N-LINKED (GLCNAC...) (POTENTIAL).
CC CARBOHYD 681 681 N-LINKED (GLCNAC...) (POTENTIAL).
CC SEQUENCE 807 AA; 309525F9EA9A99A CRC64;
CC -----
CC Query Match
CC Best Local Similarity 5.2%; Score 428.5; DB 1; Length 807;
CC Matches 133; Conservative 64; Mismatches 158; Indels 107; Gaps 27;
2Y 849 QEKVCPDPCKLSDWSSGSCSS-CGIGVIRSKWLKEKYNGRCPKLDLKNQVH 907
3b 432 EEKDDPTETCIYSNWPSACSSCTCEKGMQRMQLKAO-LDLVPCP-----DTQ 484
2Y 908 EAVPCY-----SECNOYSWVHEWSSCKINNELRLSCGGGTQSRKIRCVNTADGE 958
3b 485 DFQPCMGPGCDEGSGCTTSEWIT--WSPC-----SVSCGMGWSRE-RYVKQFPED 534
2Y 959 GGAVDNLCNQDIP-PEIOSCSL---MCPNECMSEWGLWSKCPQSCDPHTMQRTREL 1014
3b 535 G-----SVC---MLPTEETKCTVNEECSPSSCLVTEWGEWDDCSATCGMGMKRRMYK 586
2Y 1015 LRPLNSRTCAEDSQVQPCLLNCFQVNL---TEWSTCOLSENAPCGQVRT--RL 1069
3b 587 MSPADGSMCKAETSAQKMWPE-CHTIPCLLSFWSEWSDCVT---CCKMRTQRM 641
2Y 1070 SCVCSGDKPMSQCEOHNLKPEQMSIPCLVCEVNVCLSGWTAWTECSQTGCGGRMS 1129
3b 642 KSLAELG-----DCNE-DLEQAECMLP---ECPIDCELSEWSQSECKCK-GHMI 690
2Y 1130 RTFLIMPTGEGRCPCPTELTOEKTCPVTPCYSWVLGNWSACKLEGDCGEGVQIRLSLC 1189
3b 691 RTTIQMEPFGGAPCP-ETVQKCKRKAC---LRSPSIQKLWRARE----- 736
2Y 1190 MVHSGSISHAAGRVADALCGEMPFDQSILKQLCSVPCPGDCHLTSEWSTCELTICID- 1248
3b 737 -----SRSEQLREBSDEQ-----FPG-CRMRPMTAWSECTKLGCGGI 774
2Y 1249 -RSFETVGRSRSRTFTIQSFENQDSCPQVLETRPCTGK 1289
3b 775 QERYMTVKRKFKSQF-----TSCKDK-KSIRACNVHPC 807
```

RESULT 5  
FSPO\_XENLA

```
ID ID FSPO_XENLA STANDARD; PRT; 803 AA.
AC P35447;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE F-spondin precursor.
OS Xenopus laevis [African clawed frog].
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=93376785; PubMed=8367492;
RT Ruiz i Altaba A, Cox C, Jessell T.M., Klar A.;
RT "Ectopic neural expression of a floor plate marker in frog embryos
RT injected with the midline transcription factor Pintallavis.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:8288-8272(1993).
CC -!- FUNCTION: Promotes the attachment of spinal cord and sensory
CC neuron cells and the outgrowth of neurites in vitro. May
CC contribute to the growth and guidance of axons in both the spinal
CC cord and the PNS.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: Expressed at high levels in the floor plate.
CC -!- SIMILARITY: Contains 6 TSP type-1 domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L09123; AAA19105.1; -.
CC PIR; A47723; A47723.
CC InterPro; IPR002861; Reeler.
CC InterPro; IPR000884; TSP1.
CC Pfam; PF02014; Reeler; 1.
CC Pfam; PF00090; TSP1; 6.
CC SMART; SM00209; TSP1; 6.
CC PROSITE; PS50092; TSP1; 6.
CC Glycoprotein; Signal; Repeat; Cell adhesion.
CC SIGNAL 1 23 POTENTIAL.
CC CHAIN 24 803 F-SPONDIN.
CC DOMAIN 436 489 TSP TYPE-1 1.
CC DOMAIN 495 549 TSP TYPE-1 2.
CC DOMAIN 552 605 TSP TYPE-1 3.
CC DOMAIN 608 662 TSP TYPE-1 4.
CC DOMAIN 664 717 TSP TYPE-1 5.
CC DOMAIN 750 802 TSP TYPE-1 6.
CC CARBOHYD 210 210 N-LINKED (GLCNAC...) (POTENTIAL).
CC CARBOHYD 677 677 N-LINKED (GLCNAC...) (POTENTIAL).
CC SEQUENCE 803 AA; 90702 MW; D3A54E329548ABD9 CRC64;
CC -----
CC Query Match
CC Best Local Similarity 4.8%; Score 393.5; DB 1; Length 803;
CC Matches 132; Conservative 71; Mismatches 170; Indels 185; Gaps 25;
QY 760 SRKKEKQ-----DSDLYPLVETELCPDEFISQYGNW-----SDCILPGRREPHR 807
3b 403 ARKGEQCNIIIPDNVDDIVADLVTEKDDDTPTCTIYSNWPSACSSATCDKGRVQR 462
QY 808 GLRYQADSKEGEGRLFRVACSKNGRPVDPSPSCSSGVIQEKVIP-CPFD-----CKL 862
3b 463 MLKQLD-----LSVPCPTQD-----FECMGPGGSDDEASTCM 498
QY 863 SDWSWGSWSSCGSGIGVIRSKWLKXPYNGRCPKLDLKNQVHEAVPCYSECNQYSW 922
3b 499 SEWITWSPCSASCWGIEVREYVYKQFPEDG----- 529
QY 923 VEHWSCKINNELRLSCGGGTQSRKIRCVNTADGEGAVDSNLCNQDEIPEPTQSCSL- 981
```

Db 530 ---SLCKVPTPE-----ETEKGINV 545  
Qy 982 -MC-PNEUCMSEGLWSKPCOSCDPHMTMORTRHLLRPSLNSRTRCARDSDOVOPCLLNENC 1039  
Db 546 EECBPSSCIVTEWAEWECBSATCMGKKHMRMTKMPADGSMCKADTTVEVKMMP-C 604  
Qy 1040 FQFOYNL---TEWTCOLSENAVCGQVRT--RLLSVCVSDGKPVSMDOQ-EQHNLEKPKQ 1093  
Db 605 HTIPCVLSPWSEWSDCSVT---CGKGTTRQRMJK-----SPSELGDCNEELKQVE 654  
Qy 1094 RMSIPCLVECVNQLSGTAWTSCSCGCGHGSMSRTFILMPTOGEGRCPCTELTOEK 1153  
Db 655 KMLP---ECFISCLTEWSTWSCNCSCKG-GNMTIRMTLMTPEPQGGAVCP-ETVQRK 709  
Qy 1154 TCPVTPCYSVVLGNWSACKLEGGDCGEGVQIRSLSCVMVHSGSISHAAGRVEDALCGEMPF 1213  
Db 710 KC-----RLKCKQSSQNRERRHLKDKAREKRSEKIK 740  
Qy 1214 QDSILKQLCSVPCGDCHLTWSEWSTCELTCDIG--RSPETVGRQSRSTFTFIQSFENQ 1271  
Db 741 EDSGEG---YFV---CCKKPTWATTECTKFCGGGIGERFMTVKRPFKSKQF----- 786  
Qy 1272 DSCPQVLETRPCTCGKC 1289  
Db 787 TSCXDK-KEIRACNVHPC 803

RESULT 6  
AT12 HUMAN  
ID AT12 HUMAN STANDARD; PRT; 1593 AA.  
AC P58397;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DE 15-MAR-2004 (Rel. 43, Last annotation update)  
DE ADAMTS-12 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase  
DE with thrombospondin motifs 12) (ADAM-TS 12) (ADAM-TS12).  
GN ADAMTS12.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Fetal lung;  
RX MEDLINE=21264577; PubMed=11279086;  
RA Cal S. Argueelles J.M., Fernandez P.L., Lopez-Otin C.;  
RT "Identification, characterization, and intracellular processing of  
RT ADAM-TS12, a novel human disintegrin with a complex structural  
RT organization involving multiple thrombospondin-1 repeats.";  
RL J. Biol. Chem. 276:17932-17940(2001).  
CC -!- COPACITOR: Binds 1 zinc ion per subunit (By similarity).  
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular  
CC matrix (By similarity).  
CC -!- TISSUE SPECIFICITY: Expressed exclusively in fetal lung. Is widely  
CC expressed in gastric carcinomas and in cancer cells of diverse  
CC origin.  
CC -!- DOMAIN: The spacer domain and the TSP type-1 domains are important  
CC for a tight interaction with the extracellular matrix (By  
CC similarity).  
CC -!- PTM: The precursor is cleaved by a furin endopeptidase.  
CC -!- PTM: IS SUBJECTED TO AN INTRACELLULAR MATURATION PROCESS LEADING  
CC TO A FRAGMENT CONTAINING THE N-TERMINAL REGION INCLUDING THE  
CC METALLOPROTEINASE, DISINTEGRIN-LIKE, CYS-RICH AND TS-1 DOMAINS AND  
CC THE C-TERMINAL FRAGMENT CONTAINING THE SPACER 2 AND THE FOUR TS-1  
CC DOMAINS.  
CC -!- SIMILARITY: Belongs to peptidase family M128.  
CC -!- SIMILARITY: Contains 1 disintegrin-like domain.  
CC -!- SIMILARITY: Contains 1 PLAC domain.  
CC -!- SIMILARITY: Contains 8 TSP type-1 domains.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration -  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
DR EMBL; AJ250725; CAC20419.1; -.  
DR Genbank; HGNC:14605; ADAMTS12.  
DR MIM; 606184; -.  
DR InterPro; IPR001762; Disintegrin.  
DR InterPro; IPR001818; Pept\_M10A\_M12B.  
DR InterPro; IPR006025; Pept\_M\_Zn\_BS.  
DR InterPro; IPR001590; Peptidase\_M12B.  
DR InterPro; IPR002870; Peptidase\_M12B\_N.  
DR InterPro; IPR000884; TSP1.  
DR InterPro; IPR008085; TSP 1.  
DR Pfam; PF01562; Pep\_M12B\_propep; 1.  
DR Pfam; PF01421; Repolysin; 1.  
DR Pfam; PF00090; TSP 1; 6.  
DR PRINTS; PR01705; TSPREPEAT.  
DR SMART; SM00209; TSP1; 8.  
DR PROSITE; PS0215; ADAM\_MPEPO; 1.  
DR PROSITE; PS00466; CYSTEINE\_SWITCH; FALSE NEG.  
DR PROSITE; PS00427; DISINTEGRIN\_1; FALSE NEG.  
DR PROSITE; PS0214; DISINTEGRIN\_2; FALSE NEG.  
DR PROSITE; PS0300; PLAC; 1.  
DR PROSITE; PS0092; TSP1; 6.  
DR PROSITE; PS00442; ZINC\_PROTEASE; 1.  
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;  
KW Repeat; Extracellular matrix.  
FT SIGNAL 1 25 POTENTIAL.  
FT PROPEP 26 240 BY SIMILARITY.  
FT CHAIN 241 1593 ADAMTS-12.  
FT DOMAIN 241 464 METALLOPROTEASE.  
FT DOMAIN 465 544 DISINTEGRIN-LIKE.  
FT DOMAIN 542 597 TSP TYPE-1 1.  
FT DOMAIN 597 700 CYS-RICH.  
FT DOMAIN 701 826 SPACER 1.  
FT DOMAIN 823 882 TSP TYPE-1 2.  
FT DOMAIN 886 942 TSP TYPE-1 3.  
FT DOMAIN 943 996 TSP TYPE-1 4.  
FT DOMAIN 996 1315 SPACER 2.  
FT DOMAIN 1312 1365 TSP TYPE-1 5.  
FT DOMAIN 1367 1421 TSP TYPE-1 6.  
FT DOMAIN 1422 1470 TSP TYPE-1 7.  
FT DOMAIN 1471 1531 TSP TYPE-1 8.  
FT DOMAIN 1534 1574 PLAC.  
FT DOMAIN 302 305 POLY-GLU.  
FT SITE 208 208 CYSTEINE\_SWITCH (POTENTIAL).  
FT METAL 392 392 ZINC (CATALYTIC) (BY SIMILARITY).  
FT ACT\_SITE 393 393 BY SIMILARITY.  
FT METAL 396 396 ZINC (CATALYTIC) (BY SIMILARITY).  
FT METAL 402 402 ZINC (CATALYTIC) (BY SIMILARITY).  
FT CARBOHYD 105 105 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 125 125 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 215 215 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 485 485 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 685 685 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 790 790 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 951 951 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1104 1104 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1275 1275 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1300 1300 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1320 1320 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1371 1371 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1378 1378 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1503 1503 N-LINKED (GLCNAC. .) (POTENTIAL).  
SQ SEQUENCE 1593 AA; 177545 MW; 07F9F48363BD83A3 CRC64;

Query Match 4.4%; Score 360; DB 1; Length 1593;  
Best Local Similarity 19.7%; Pred. No. 2.8e-16;  
Matches 249; Conservative 114; Mismatches 388; Indels 512; Gaps 62;

```
2Y 46 PMSNCKGCKGLQHRTRAVIAP-PLFGGLQCP-----NLTESBAC--DAP-----ISC- 91
2b 550 PWSHCRTCGAGVQSAERLCNNPEKFGKYGKGTGERKRVLCNVHPCRSEAFTRQMQCS 609
2Y 92 -----PLGEEVTFSLKVPKSWKCRPLHLKEIN-----PSGRTV- 125
2b 610 EFTVYKVELHWFFIFPAHPCEL-YCRPIDGOFSEKXLDVADIGTFCFFGGNSRNV 668
2Y 126 -----LDFNSDSNE-----RVTFKHQSYKAHHHSKWSAIEIGY 158
2b 669 INGIKRWGCDYEDSNATEDRCGVCLGSSCQTVRKMFKEKGGY-----VDIGL 721
2Y 159 ---QTRQVSCSTRSDGNAMLSLCLQ-----181
2b 722 IPKGARDIRVMEIEGAGNFAIRSEDEPKYIINGGFIOWNGNYKLAVFOYDRKGBLE 781
2Y 182 -----SFPLTVOSC:MPKDCETS---QWSSWSPCKSKTCKS 213
2b 782 KLWATGPTNESVWIIQLFQVTFPIKYEYTIQKGLDNDVEQMYFWQYGHWTCECVTGT 841
2Y 214 GLLLPGRSRNVKH-----MAIGGKECPLELLEKEACIVEGELLQOQPR 259
2b 842 GIR---ROTAHCKKGRGVKATFCDPETQPNROKKCHE-----KACPP 883
2Y 260 YSWRTSEWKECQVSLLEQDDPHWHVTFVCGGGIQTREVYCAQSVPAAPAAALRAKEVSEP 319
2b 884 -RWAGENEACSAIC---GPH-----GEKRTVLICIQM-----VSDQALPP 922
2Y 320 VEKALCVGAPLPSQJCN-IFCSTDCIYSSWSAMGLCIHENCHEPQGGKGFRTQRHVL 377
2b 923 TDCQHLKPKTLLS---CNRDILCPSDWTGVNNS-----ECVSVCQ-GGVRIK- 966
2Y 378 MESTGPAGCHPLVESVPC---EDPMC---YRWLASGICFDPHGKGLHRLIKAVQOND 432
2b 967 -----SVTCAKHDEPCDVTXKPNRGCALC---GLQCPSPSRVLKPC---N 1005
2Y 433 RGEDVSG-----SILCPVPPPPER-KSEICPRMDCVLSEWTESSCSQSCNKNSD--GKQ 485
2b 1006 KGTISNGKPNPILKVPVPTSPRMLTTTGESMSTSTPAISSPPTASKEGLDGKQ 1065
2Y 486 TR-----SRTILAGEGKPCPPSQAAL-----QEHRLNDHSCMLHWTSPWGPCS 533
2b 1066 WQDSSTQPELSRRLYSTGSTSQPILTSQSLSIQFSENV-----SSSDTGPTS 1114
2Y 534 EDTLVATNATIGNGEATCGVGIOTRR-----VFCVKSHVQC----- 571
2b 1115 EGLVAT-----TTSGLGSSRNPIWTPVTPVNTLTKPEMEIHSGEEREQ 1164
2Y 572 -----VMTK-RCPDSTRP--ETVAPCFPLCKKDCIVTAFSEWTPC----- 608
2b 1165 PEDKDBSNPVIWTKIRVGNDAFVSESTEMPLAPLTPD--LGSRESWPPFSTVMEGLLPS 1222
2Y 609 --PRMCQAGNATVKSRYLIIICEAN-----GGQECPTLYEERECEDEVSLCPVYRWKP 661
2b 1223 QRTTSETGTPRVE---GMVTEKPANTLPLLGQHQ----- 1256
2Y 662 QKWSPCILVPESWQIGTSSSEACGGLQTRAVSICISDDNRAEMWECILKOTNGMFLVQ 721
2b 1257 -----EPGSGKTANRNHLKLPNNNNQTKSEPVLTEDATSLITE 1295
2Y 722 ECTVPCREDCTFT-----ANSKTTPTSTNCEATKSRRLQITGSKRKEKC---Q 767
2b 1296 GFULNASNYKQLTNGHGSARHIVGNWSE---CSTTCG-----LGAYWKRVCCTTQM 1343
2Y 768 DSDLY-----PLVETELCPCEFTISQPYGNNSDCILPEGREPHRGLRVQADSKEGEG 821
2b 1344 DSDCAAIQRPDPAPKACHLAPC---AGWKVGNWSK-----SNCSGG 1382
2Y 822 LRPRVACSD-KNGRPVDPFSFSSSYIIEKCVIPC-PFDCKLSDWSSWGSCSSCGIGV 879
2b 1383 FKIREIQCVDSRDLRNLRFPHCFQFLAGIIPPLSMSCNPEPCFAWQVWPSQCSRSRGGGV 1442
2Y 880 RIRS-----KWLKEKPVNGRGPCKFLDLKQNVHEAVPCYSECNQY---SWVVEHWS 928
```

```
Db 1443 QERGVPFGGLCDWTK-----RPTSTM-----SCNEHLCCHWATGNWL 1481
QY 929 KXINNELRLRCGGTQSRKIRCVNTADGEGGAVDSNLCNQDEIPBETQSCS----- 980
Db 1482 C-----SYSCGGGFKRIVQCVPSGKNTBDOOQCLDKHPRPFKCNQOACKSA 1534
QY 981 -LMCPN-----ECVMSEWGLMSKCPQSCDPTMORRTHLL 1015
Db 1535 DLLCTKDLKLSASFCQTLKAMKCSVTVRAECCFS-----CPQTHITHTQRQRRL 1587
QY 1016 RPS 1018
Db 1588 QKS 1590

RESULT 7
SM5B_MOUSE
ID_SM5B_MOUSE STANDARD; PRT; 1093 AA.
AC Q60519;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Semaphorin 5B precursor (Semaphorin G) (Sema G).
GN SEMA5B OR SEMAG OR SEMG.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=NMRI.
RX MEDLINE=96414430; PubMed=8817451;
RA Adams R.H., Betz H., Puschel A.W.;
RT "A novel class of murine semaphorins with homology to thrombospondin
RT is differentially expressed during early embryogenesis.";
RL Mech. Dev. 57:33-45(1996).
CC -!- FUNCTION: May act as positive axonal guidance cues.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: In adult, only detected in brain.
CC -!- DEVELOPMENTAL STAGE: Differentially expressed in embryonic and
CC adult tissues. Its abundance decreases from E10 to birth.
CC -!- SIMILARITY: Belongs to the semaphorin family.
CC -!- SIMILARITY: Contains 1 Sema domain.
CC -!- SIMILARITY: Contains 7 TSP type-1 domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X97818; CAA66398.1; -.
CC MGD; MGI:107555; Sema5b.
CC InterPro; IPR003659; Plexin-like.
CC InterPro; IPR002165; plexin_repeat.
CC InterPro; IPR001627; Sema.
CC InterPro; IPR000884; TSP1.
CC InterPro; IPR008085; TSP_1.
CC Pfam; PF01437; PSI; 1.
CC Pfam; PF01403; Sema; 1.
CC Pfam; PF00090; tsp_1; 5.
CC PRINTS; PR01705; TSP1REPEAT.
CC SMART; SM00423; PSI; 1.
CC SMART; SM00630; Sema; 1.
CC SMART; SM00209; TSP1; 4.
CC PROSITE; PS00092; TSP1; 5.
KW Signal; Transmembrane; Repeat; Multigene family; Neurogenesis;
KW Developmental protein; Glycoprotein.
FT SIGNAL 1 19
FT CHAIN 20 1093 SEMAPHORIN 5B.
```

FT DOMAIN 20 978 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 979 999 POTENTIAL.  
FT DOMAIN 1000 1093 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 336 518 SEMA.  
FT DOMAIN 551 605 TSP TYPE-1 1.  
FT DOMAIN 606 662 TSP TYPE-1 2.  
FT DOMAIN 664 713 TSP TYPE-1 3.  
FT DOMAIN 721 776 TSP TYPE-1 4.  
FT DOMAIN 795 850 TSP TYPE-1 5.  
FT DOMAIN 852 907 TSP TYPE-1 6.  
FT DOMAIN 908 952 TSP TYPE-1 7.  
FT CARBOHYD 59 95 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 95 95 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 157 157 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 287 287 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 378 378 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 539 539 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 547 547 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 602 602 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 728 728 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 944 944 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 1093 AA; 120326 MW; 29E5C9B1E8108717 CRC64;  
Query Match 4.4%; Score 358.5; DB 1; Length 1093;  
Best Local Similarity 22.4%; Pred. No. 2.3e-16;  
Matches 180; Conservative 66; Mismatches 248; Indels 309; Gaps 41;  
QY 579 DSTRPETVRPFLPCKKDCIVTAFSEWTPCPRMCOAGNATVKQRYRIIOEAAAGGQEC 638  
DB 395 EAVQVPTTEPC-----VTQDSVRFSLVDLV---QA 423  
QY 639 PDLTVE-----REC--EDVSLCPVYRWKPKQKSPCLIVPSVW 675  
DB 424 KDTLHVLYIGTESITLKALSTASRSRGVYBELHVLPPGRLEPLRSILHSARALF 483  
QY 676 QGITG-----SSEAC-----GKGLQTRAVSCISDDNSAEWECLKOTNMPLLVQECT 724  
DB 484 VLSRVRIRIPLERSAVHSQACLGADDFCGWDGK--QLCTLEDSSNWSLWQNIIT 541  
QY 725 V-PCR---EDCTFTAWSKFTPOSTNCEATKRRRLQTKSRKKEKQSDLYPLVETELC 780  
DB 542 TCPVRNVRDGGFGPSPWKPC-----EHLDDGNSGSLCRARS-----580  
QY 781 PDEFISQPYGNWSDCIIPEGREPHRGLRVQADSKCEGELRPAVACSKRGPVDPDS 840  
DB 581 -CDS-----PRPR-----CG-----GLECL---GFSIHIA 601  
QY 841 FCSGGYIQEKCVIPCPDFCKLSDSSWGSSCSGIGVIRIRSKWLKE-KPYNGRCP-P 898  
DB 602 NCSRNG-----AWTAWSSWAQCSSTSGIGVQVRQSRCSNPAPRHGRICVG 647  
QY 899 KLDLKNQVHEAVPCYSECNQVSWVVEHWSCKINNELSLRCGGTQGRKRCVNTADGE 958  
DB 648 KREERFCNENTPCVPVIFWASW--GSWSKCSNN-----CGGVQSRRRSCENGNSCP 698  
QY 959 GGAVDSNLCNDEIPEPTQSCSLMCPNECVMSWGLW-----SKCPQSC---1002  
DB 699 GGVFEKTCN-----PEA-----CPEVRNTPWTPNLPVAVTQGGARQORFRTCTAP 747  
QY 1003 --DPHTMQ--RRTHLLRPSLNSRTCAEDSOVQCLLNENCFQYFN-----ITEWSTC 1052  
DB 748 LPDPHGLQFGKERTETRTCPADGTGACDLDALVEDLLRSGSTSPHTLNGWATWGPWSSC 807  
QY 1053 QLSENAPCGQVTRTLLSCVSDGDFVMSQCEQENLEKPKQRMSTPCLVE-----1102  
DB 808 -----SRDELGLPRVRRKRTCT-----NP-----EPNNGGLFCVGDAAETQDCNPQ 847  
QY 1103 -CVVNCQLSGMTAWTECSQTQGHGRMSRTRFIIMPTQGEGRP-CPTBELTOEKTCPTVPC 1160  
DB 848 ACPVRGAWSCMTAWSCSASCG-GGHYQTRTSCTSPAPSPGEDIICLGLHTEALCSQAC 906

QY 1161 -YSVVLGNWSACKLRGGDCGE-GVQIRSLSCMVHSGSISHAAGRVEDALCGEMPPQDSIL 1218  
DB 907 PEGMSL--WS-----EWGVTEDGAQSRSRSC-----EELL-----935  
QY 1219 KOLCSVPCCPGDCHLTSEWSEWSTCELTCDIGRSFETVGRQSRRTFIQSFPENQDSCP---1275  
DB 936 -----PGFGAC-----VGNSSQSR-----PCPYSE 955  
QY 1276 -----QQVLETRPCTGGKCYH 1291  
DB 956 IPVILPASSVEITSCGGFNLIH 978  
RESULT 8  
BAR3\_CHITE STANDARD; PRT; 1700 AA.  
ID BAR3\_CHITE  
AC Q03376;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Balbiani ring protein 3 precursor.  
OS BR3.  
OS Chironomus tentans (Midge).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Nematocera; Chironomoidea;  
OC Chironomidae; Chironominae; Chironomus.  
OC NCBI\_TaxID=7153;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Salivary gland;  
RX MEDLINE=90172404; PubMed=1689777;  
RA Paulason G., Lendahl U., Galli J., Ericsson C., Wieslander L.;  
RT "The Balbiani ring 3 gene in Chironomus tentans has a diverged  
repetitive structure split by many introns."  
RL J. Mol. Biol. 211:331-349 (1990).  
CC -!- FUNCTION: Used by the larvae to construct a supramolecular  
role as a transport protein that binds to other proteins  
intracellularly and in the gland lumen in order to prevent these  
from forming water-insoluble fibers too early.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- TISSUE SPECIFICITY: Salivary gland.  
CC -!- DOMAIN: Has 82 approximate repeats of Cys-x-Cys-x-cys.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
DB EMBL; X52263; CAA36506.1; --  
DR PIR; S08167; S08167.  
DR HSSP; P15358; 1SKZ.  
DR InterPro; IPR004153; CXXC-repeat.  
DR Pfam; PF03128; CXXC; 71.  
DR Repeat; Signal.  
FT SIGNAL 1 20 POTENTIAL.  
FT CHAIN 21 1700 BALBIANI RING PROTEIN 3.  
SQ SEQUENCE. 1700 AA; 186145 MW; 34202B28521B0815 CRC64;  
Query Match 4.3%; Score 357.5; DB 1; Length 1700;  
Best Local Similarity 19.6%; Pred. No. 4.5e-16;  
Matches 303; Conservative 146; Mismatches 595; Indels 499; Gaps 89;  
QY 28 EQACLIPCR--DCVVSEFLPMSNCKGCKLQHRTRAVIAPPLFGGLQCPLNLTESRAC 85  
DB 127 EKSCACVCFNAKCTAPQVNNKDTCCCGCPVNNQBPADGCTKPLIWDKVDG-----RC 179  
QY 86 DAPISCPGLGEETFTSLKVGPMW-----KCRPLHLKEINP 120

Db 180 ECPFKKCGKNNR-----DWSDSKSCCKGDKGKQSGKINWKNKNCRCICP---TAPF 228  
Qy 121 SGRVLDNFSDSNE-----RVTFHQYKAHHHSKSW---AIEIGYQTRVOSTRSDGON 172  
Db 229 AGGCSAPLKWDDDKSCACAPAKMEKKEKCVESKINWPNPTCEG--CAQLNCPDNKKAN 286  
Qy 173 AMLSLCLQDSFPLTVQSCIMPDKCETSQNSWSFPCS-----KTCRSGSLLPGRFSR 223  
Db 287 KETQCQ-----ECKEVKNGGQVFCCKDSCSCVCFGGDKDKTCTAPQYDYG---332  
Qy 224 SRNVKHAIGGKECPLEKEACIVEGELLQOCPR--YSWRTSEWK--ECOVSLLELLEQQDP 281  
Db 333 -----VACSCSPVNMKPA-----DCPRPKQWDKECECECPVK-----368  
Qy 282 H-----WHVT-----GPVCGGGIOTREYVCAQSVAPAAALRAKEVSPVEKALC 325  
Db 369 HDCKNGKVMDETTCQCTCPDAPVCTAGKER-----CGES-----CECKCINREPKEG-C 417  
Qy 326 VGPAPLPSQLCNIPCSDDCIIVSSWSAWG-----LCIHENCHEPQGGKGRTEQRHVLM 379  
Db 418 AKPLVWNENTCKVCVPADKQVSGGSGKSFNKLTCQCECDQASGKGLRWV-----471  
Qy 380 STGPAGHCPHLVSPCEDPMVCYRWLASBGIC-----FPDHGKCGLGHRIKAVQCONDRG 434  
Db 472 -----ADTKCE-----CQGMPEEG--CGKQTWISDKCKCECSPTI---TCQAPQI 513  
Qy 435 EDVS--GSLCPVPPPERKSCIEPCRMDCVLSEWTEWSSCSQSCSN--KNSDGKOT--RSR 489  
Db 514 LDLLNTECKCPVNMVLAQEKCKGP-----RQWTD--SKLCECSTTPATCEGCKQWCEB 565  
Qy 490 TILALAGEGKCPPPSQAQO-----EHLRNDHSCQWLHWETSP--WG-----PCSSEDTL 537  
Db 566 ACQICFGGKNGKNGKFFDKPSCECKCKNPTC-----TSQVWDADDCECKCFPKDQ 619  
Qy 538 VTALNATIG--WNGEATCGVIGIOTRRVFCVSHVQVMTKCPDSTDPTRVTRPFCFLPC--593  
Db 620 KPGCGDGGGOKWN--DRVCSGCGPVRPDDTNGQIYNIT-----CAGCGGI 664  
Qy 594 -KKDCIVTAFSEWTPCPRMCOAGNATVKQSRRIIIOEAAANGGOEGCPDTLYEBERECEV 652  
Db 665 DKPSCPQOIIYNWKTDCBENG-----MKEPVGG--CGATWLDDEQ---706  
Qy 653 LCPYRWKPKWPKSPCLILVPSVWQGTGSSEACGKGLQTRAV-----694  
Db 707 -----CDCVPKPGGGTGAQKWDKTKCKCEKEMPTGGCENKWKCDET 752  
Qy 695 -----SCISDDNPSAEMWEL-----KQTHGMPILVQECTVPCREDCTFTWASKFTF 741  
Db 753 CDCVPQKNTCIAPKVDATKSCICVNPFPKNSPQVLKDTCCCGCNVSKCAPOKFIE 812  
Qy 742 CSTNCEATKRRRLQTKSRKKEKCDSDILPLVETELCPDEFISQYXGNWSDCLPEG 801  
Db 813 NIDCACP-----NKKQCK-----APLWS-----DEF-----CDCVCPN- 842  
Qy 802 RREPHRGLRVQADSKEGEGRLPRAVACS--DKNGRPVDSFC--SSSGYIOBKCVIFCP--857  
Db 843 -----SAGMKTCLSPKWNKVTCTCDCH--PPKPDCCPGTKQWMDKCKCPCNA 890  
Qy 858 -FDC-----KLSDDSWSGSCSSSGIG---VRIKSKWLKXPYNG---GRPCPKLLDKQ 905  
Db 891 QTDGAGGKENDF-----TCSGCGPSGLKLDCTGNTKWSAETCTCGGVNRCG--NLKNF 944  
Qy 906 VHEAVPCYSCNQSVVVEHWSCKINNELSRJRCGGTQSRKIRCVNTADGEGGAVDSN 965  
Db 945 NDNL--COCECKN-----KQBMANCKSPRTWNYDTC-----KVCCKN-----ADSD 984  
Qy 966 LQNDIEPPTQSCSLMCPNECVNWSGLWSKCPQ-----SCD-----PHTM 1007  
Db 985 DCVKPQIWLDDQ--CKCGCPASAQWT-----CPANKRFIEKSCCEKSPSPFPQCK 1036  
Qy 1008 ORTRHLLRSLNSRTCAE-----DSQVQPL---LNENCFQFQYNLTWSTCQLSENAP 1059  
Db 1037 KWNEDKCVVECANVTKCEGFORWCDNCK--CICPQVNTKCSDKQKFIKCEGCGCBTQ 1095

## RESULT 9

SMSA MOUSE  
ID SMSA MOUSE STANDARD; PRT; 1077 AA.  
AC Q62217;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Semaphorin 5A precursor (Semaphorin F) (Sema F).  
GN SEMA5A OR SEMAF OR SEMF.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NMRI;  
RX MEDLINE=96414430; PubMed=8817451;  
RA Adams R.H., Betz H., Puschel A.W.;  
RT "A novel class of murine semaphorins with homology to thrombospondin  
is differentially expressed during early embryogenesis.";  
RL Mech. Dev. 57:33-45(1996).  
CC -!- FUNCTION: May act as positive axonal guidance cues.  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -!- TISSUE SPECIFICITY: IN ADULT, DETECTED IN LIVER, BRAIN, KIDNEY,  
CC HEART, LUNG AND SPLEEN.  
CC -!- DEVELOPMENTAL STAGE: DIFFERENTIALLY EXPRESSED IN EMBRYONIC AND  
CC ADULT TISSUES. ITS ABUNDANCE DECREASES FROM E10 TO BIRTH.  
CC -!- SIMILARITY: Belongs to the semaphorin family.  
CC -!- SIMILARITY: Contains 1 Sema domain.  
CC -!- SIMILARITY: Contains 7 TSP type-1 domains.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X97817; CAA66397.1; ..  
DR MGD; MGI:107556; Sema5a.  
DR GO; GO:0016021; C:integral to membrane; IDA.  
DR GO; GO:0008046; F:axon guidance receptor activity; IDA.  
DR GO; GO:0007411; P:axon guidance; IMP.  
DR InterPro; IPR003659; Plexin-like.  
DR InterPro; IPR002165; Plexin\_repeat.  
DR InterPro; IPR001627; Sema.





```
DR SMART; SM00209; TSP1; 6.  
DR PROSITE; PS0092; TSP1; 6.  
KW Signal; Transmembrane; Repeat; Multigene family; Neurogenesis;  
KW Developmental protein; Glycoprotein.  
FT SIGNAL 1 22  
FT CHAIN 23 1074  
FT DOMAIN 23 968  
FT TRANSMEM 969 989  
FT DOMAIN 990 1074  
FT DOMAIN 226 507  
FT DOMAIN 540 593  
FT DOMAIN 595 651  
FT DOMAIN 653 702  
FT DOMAIN 707 765  
FT DOMAIN 784 839  
FT DOMAIN 841 896  
FT DOMAIN 897 944  
FT CARBOHYD 142 142  
FT CARBOHYD 168 168  
FT CARBOHYD 227 227  
FT CARBOHYD 277 277  
FT CARBOHYD 323 323  
FT CARBOHYD 367 367  
FT CARBOHYD 437 437  
FT CARBOHYD 536 536  
FT CARBOHYD 591 591  
FT CARBOHYD 717 717  
FT CARBOHYD 933 933  
FT CONFLICT 56 56  
FT CONFLICT 149 149  
FT CONFLICT 382 382  
FT CONFLICT 494 494  
SQ SEQUENCE 1074 AA; 120570 MW; EE3DB763CBE29407 CRC64;  
  
Query Match 4.2%; Score 345.5; DB 1; Length 1074;  
Best Local Similarity 22.6%; Pred. No. 1.7e-15;  
Matches 141; Conservative 59; Mismatches 168; Indels 255; Gaps 30;  
  
QY 794 SDCIL-----PEGRRPHRLRVOADSKCGSLRFRAY----- 827  
DB 441 SCLLEBIEELFPERREPIRLSLQLHSQVLFVGLRHVVKFLKRCQFVTRSTCIGAQ 500  
QY 828 -----ACSDKN----- 827  
DB 501 DPYCGDWMVKKCTSLSESLMTQWEOISACPTRLNLTVDGHGVMSPTPCPTHGSAV 560  
QY 838 DPSCSSSGVIOEKVPCP-----FDCK-----LSDHSSWGSCSSSGIGV 879  
DB 561 GSCLCRT-----RSCDSPAQCGGWQCEGPMETIANCSRNGWTPTWTSWSPCTTCGIGF 615  
QY 880 RIRSKWLKE-KPVNGGRPCPKLDLKNQVHBAVPCYSECNOYSWVVEH-----WSSCK 930  
DB 616 QVRQSCSNPTPRHGRVVCVQGNREERY-----CNEHLCPHFMTWGWGPHERC- 665  
QY 931 INNELSLRCGGTQSKIRCVNTADEGAVDSNLNQDEIPPEQSCSL-MCPNECV 989  
DB 666 -----TAQCGGSIQARRICENGPDGAG-----CN-----VEYQSCNTNCPCLKKT 707  
QY 990 SEWGLSKCQSCDPHMQWRTTH-----LL-----RPSLNSRTCAEDSQVQCLN 1036  
DB 708 TPWTPTWPNVINDSGHDHVEQFRYTCARLADPNLLEVGQRTEMRYCSDG-TSCGSTD 766  
QY 1037 ENCFQF-----OYNL-----TENSTCOLSENAPCGGVTRLLSCVCSGPK-VSM 1081  
DB 767 GLSGDFLRAGYSAAHTVNGAWSAWTSWQC-----SRDCSRGIENR--KRVCNNPEPKYGG 820  
QY 1082 DQCEQHNLEXPQMSIPCLVECVNQLSGWTATWESQTCGCGGRMSRTRFLIMTQGE 1141  
DB 821 MCLGPLELEQECNLTCPVDGWSC-----WSPWTKSATCG-GGHVMTVSCSNP- 871  
QY 1142 GRPCPTLTQKTCPTVTPCYSWVLGNWSACKLEGDCGEGVQIRSLSCMVHSGSISHAAG 1201  
DB 872 -----APAY-----GGDICLGL----- 883
```

```
QY 1202 RVEDALCGEMPQDSILKQLGVPCGCHLTSEWSTCETCIDGRSFETVQGRSR 1261  
DB 884 HTEALCN-----TQCPES--WSESDWSECEAS-----GVQVRR 918  
QY 1262 TFIQSFENQDSFCQVLETRPC 1284  
DB 919 QCIL-LFFPMGSCGSGNTTESRPC 940  
  
RESULT 11  
ID SMSB_HUMAN STANDARD; PRT; 1093 AA.  
AC Q9P283;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Semaphorin 5B precursor.  
GN SEMA5B OR KIAA1445.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=20277482; PubMed=10819331;  
RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. XVII.  
RT The complete sequences of 100 new cDNA clones from brain which code  
RT for large proteins in vitro."  
RL DNA Res 7;143-150(2000).  
CC -!- FUNCTION: May act as positive axonal guidance cues (By  
CC similarity).  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -!- SIMILARITY: Belongs to the semaphorin family.  
CC -!- SIMILARITY: Contains 1 Sema domain.  
CC -!- SIMILARITY: Contains 7 TSP type-1 domains.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL; AB040878; BAA95969.1; ALT_INIT.  
CC Genew; HGNC:10737; SEMA5B.  
CC InterPro; IPR003659; Plexin-like.  
CC InterPro; IPR002165; Plexin_repeat.  
CC InterPro; IPR001627; Sema.  
CC InterPro; IPR000884; TSP1.  
CC InterPro; IPR008885; TSP_1.  
CC Pfam; PF01437; PSI; 1.  
CC Pfam; PF01403; Sema; 1.  
CC Pfam; PF00090; tsp_1; 5.  
CC PRINTS; PR01705; TSP1REPEAT.  
CC SMART; SM00423; PSI; 1.  
CC SMART; SM00630; Sema; 1.  
CC SMART; SM00209; TSP1; 4.  
CC PROSITE; PS50092; TSP1; 5.  
KW Signal; Transmembrane; Repeat; Multigene family; Neurogenesis;  
KW Developmental protein; Glycoprotein.  
FT SIGNAL 1 26  
FT CHAIN 27 1093  
FT DOMAIN 20 978  
FT TRANSMEM 979 999  
FT DOMAIN 1000 1093  
FT CYTOPLASMIC (POTENTIAL).  
FT SEMA.  
FT DOMAIN 551 605  
FT TSP TYPE-1 1.  
FT TSP TYPE-1 2.  
FT TSP TYPE-1 3.  
FT TSP TYPE-1 3.
```

FT DOMAIN 721 776 TSP TYPE-1 4. 776 1093 AA; 119866 MW; F1PDEFB87CEAF0EF CRC64;  
FT DOMAIN 795 850 TSP TYPE-1 5. 850 1093 AA; 119866 MW; F1PDEFB87CEAF0EF CRC64;  
FT DOMAIN 852 907 TSP TYPE-1 6. 907 1093 AA; 119866 MW; F1PDEFB87CEAF0EF CRC64;  
FT DOMAIN 908 952 TSP TYPE-1 7. 952 1093 AA; 119866 MW; F1PDEFB87CEAF0EF CRC64;  
FT CARBOHYD 59 59 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 95 95 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 157 157 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 287 287 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 378 378 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 539 539 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 547 547 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 602 602 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 728 728 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 944 944 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 1093 AA; 119866 MW; F1PDEFB87CEAF0EF CRC64;  
Query Match 4.1%; Score 336.5; DB 1; Length 1093;  
Best Local Similarity 21.8%; Pred. No. 6.9e-15;  
Matches 169; Conservative 70; Mismatches 218; Indels 319; Gaps 40;  
QY 657 YRWKPKQ-KWSP-----CILVPESVWQGITGSSEACGKGLQTRAVSCISDDNRGAEM 706  
DB 347 YQENPRAWLPIANPIPNFQCGTIFE-----TGPNE-----NLTERS-----LQDAQRFLFM 393  
QY 707 MECLQKQNGMLLVOECT-----VPCREDCTFTAMSKFTPCSTNCEATKYSRRRLGTG 758  
DB 394 SEAVQVTPPEPCVTDQSVRFSLVVDLVQAKDTLHYLYIGTSGTILKALSTARSLSHG 453  
QY 759 KSRKXKQDSDPLVVELTCLPCDEFTSQPVGNWSDCILPEGRPERHRLGRVQADSKK 818  
DB 454 -----CYLSELH-----VLPGRPERHRLGRVQADSKK 482  
QY 819 EGELR-----FRAY-ACSDXNKRVPDPFCSSGYIQEK- 852  
DB 483 FVLGDLGVLRLPRLERCAAYRSQACGLGAR-----DP-YCGWDGK-QQRCSTLEDSSNNLSLW 536  
QY 853 ---VLPCEP-----FDC--- 860  
DB 537 TONITACPVRNTRDGGFGPNSWPQCHLDGDNSSGCLCARSCDSRPRCGGLDCLGP 596  
QY 861 -----KLSDWWSGSSSCGIVRIRSKWLKE-RFYNGGRPC-KPLDLKXQVH 907  
DB 597 ALHIANCSRGAWTPWSSWALCSTCGIFGQVQRSCSNPAPRHGGR-CVCKSRERFCN 656  
QY 908 EAVPCYSECNOYSVVHEWSSCKINNELRSRCGGTQSRKIRCVNTADGEGGAVDSNLC 967  
DB 657 ENTCPVPPIFWASW--GWSKCSN-----CGGMSQSRACENGNSCLGCGVEFTC 707  
QY 968 NODEIPPETQSCSLMCPNECVMSBGLW-----SKPQSCDPHTMQ 1008  
DB 708 N-----PEG-----CPEVRNTPTWPLPVNTQGGARQORFFTCRAPLA-DPHGLQ 755  
QY 1009 ---RTRHLLRLSLNRTCAEDSQVQPCLLNENCFOFN-----LTSWSTCOLSENAFC 1060  
DB 756 FGRRTETRTCPADGSGCDTALVELLRSSTSPHTVSGVAAWGPWSSC-----SRDC 811  
QY 1061 GGVRTLLSCVCSGDKGKPVSMQCEHNLKPKRMSIPCLVE-----CVANCOL 1109  
DB 812 ELGFAVRKKTCT-----NP-----EPNGLFCVGDAAEYQCNFCAPVRGAW 855  
QY 1110 SGWTAWTECQFCGGRMSRTRFIMPTQGBRCPCTELTQEKTCPTVPCYSWVLGNWS 1169  
DB 856 SCWTSWSPSCASCG-GGHYQTRTSCTSPA-----PSP----- 886  
QY 1170 ACKLEGGDCGEGVQIRLSLSCM-VHSGSISHAAGRVEDALCGEMPFDQSLKQLGCVPCPG 1228  
DB 887 -----GEDI-----CLGHT-----EELUC-----ATQACPE 908  
QY 1229 DCHLTWSEWSTCELTCDIGRSFETVGRQSRRTFIQSFENQDSCPOQVLETRPC 1284

Db 909 G--WSPWSEWSKC-----TDDGAQSRSR-HCELLPSSACAGNSQSRPC 951  
RESULT 12  
ZAN\_MOUSE STANDARD; PRT; 5376 AA.  
ID ZAN\_MOUSE O08647;  
AC O08647; O08647;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Zonadhesin precursor.  
GN ZAN.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RX MEDLINE=98123114; PubMed=9452463;  
RA Gao Z., Garbers D.L.;  
RT "Species diversity in the structure of zonadhesin, a sperm-specific  
RT membrane protein containing multiple cell adhesion molecule-like  
RT domains";  
RL J. Biol. Chem. 273:3415-3421(1998).  
RN [2]  
RP SEQUENCE OF 4864-5376 FROM N.A.  
RC TISSUE=Testis;  
RX MEDLINE=97271566; PubMed=9126492;  
RA Gao Z., Harumi T., Garbers D.L.;  
RT "Chromosome localization of the mouse zonadhesin gene and the human  
RT zonadhesin gene (ZAN).";  
RL Genomics 41:113-122(1997).  
CC -!- FUNCTION: Binds in a species-specific manner to the zona pellucida  
CC of the egg. May be involved in gamete recognition and/or  
CC signaling.  
CC -!- SUBUNIT: Probably forms covalent oligomers.  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein, exclusively on the  
CC apical region of the sperm head.  
CC -!- TISSUE SPECIFICITY: In testis, primarily in haploid spermatids.  
CC -!- DOMAIN: The MAM domains probably mediates sperm adhesion to the  
CC zona pellucida.  
CC -!- DOMAIN: During sperm migration through the reproductive tracts,  
CC the mucin-like domain might inhibit inappropriate trapping of  
CC spermatozoa or promoting adhesion to the oviductal isthmus.  
CC -!- DOMAIN: The VWFD domain 2 may mediate covalent oligomerization (by  
CC similarity to human intestinal mucin MUC2).  
CC -!- SIMILARITY: Contains 3 MAM domains.  
CC -!- SIMILARITY: Contains 25 VWFD domains.  
CC -!- SIMILARITY: Contains 1 EGF-like domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
EMBL; U97068; AAC36680.1; -;  
DR EMBL; U83190; AAC53125.1; -;  
DR PIR; T42215; T42215.  
DR MGI; MGI106656; zan.  
DR InterPro; IPR006209; EGF\_like.  
DR InterPro; IPR003645; FOIN.  
DR InterPro; IPR006210; IEFG.  
DR InterPro; IPR000998; MAM domain.  
DR InterPro; IPR002919; TIL\_Cysrich.  
DR InterPro; IPR003328; TIL\_Cysrich.  
DR InterPro; IPR001007; VWF\_C.  
DR InterPro; IPR001846; VWF\_D.  
DR Pfam; PF00629; MAM; 3.  
DR Pfam; PF01826; TIL; 25.  
DR

```

DR PFam; PF02345; TILs; 25.
DR PFam; PF00094; vwd; 4.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00274; FOLN; 11.
DR SMART; SM00137; MAM; 2.
DR SMART; SM00214; VNC; 17.
DR SMART; SM00216; VWD; 4.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 18.
DR PROSITE; PS00026; EGF_3; 1.
DR PROSITE; PS00740; MAM_1; FALSE_NEG.
DR PROSITE; PS00060; MAM_2; 3.
KW Signal; Glycoprotein; Transmembrane; EGF-like domain; Cell adhesion;
KW Repeat.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 5376 ZONADHESIN.
FT DOMAIN 18 5310 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 5311 5337 POTENTIAL.
FT DOMAIN 5338 5376 CYTOPLASMIC (POTENTIAL).
FT MAM 1. 45 210
FT DOMAIN 215 374
FT MAM 2. 377 542
FT DOMAIN 547 1170
FT DOMAIN 1171 1280
FT DOMAIN 1281 1669
FT DOMAIN 1670 2056
FT DOMAIN 2057 2459
FT DOMAIN 2460 2579
FT DOMAIN 2580 2699
FT DOMAIN 2700 2819
FT DOMAIN 2820 2939
FT DOMAIN 2940 3059
FT DOMAIN 3060 3179
FT DOMAIN 3180 3299
FT DOMAIN 3300 3416
FT DOMAIN 3417 3536
FT DOMAIN 3537 3656
FT DOMAIN 3657 3776
FT DOMAIN 3777 3892
FT DOMAIN 3893 4028
FT DOMAIN 4029 4148
FT DOMAIN 4149 4263
FT DOMAIN 4264 4383
FT DOMAIN 4384 4503
FT DOMAIN 4504 4623
FT DOMAIN 4624 4743
FT DOMAIN 4744 4863
FT DOMAIN 4864 5261
FT DOMAIN 5259 5295
FT DISULFID 5263 5274
FT DISULFID 5268 5283
FT DISULFID 5285 5294
FT CARBOHYD 339 339
FT CARBOHYD 499 499
FT CARBOHYD 1216 1216
FT CARBOHYD 1239 1239
FT CARBOHYD 1314 1314
FT CARBOHYD 1814 1814
FT CARBOHYD 1908 1908
FT CARBOHYD 1933 1933
FT CARBOHYD 2029 2028
FT CARBOHYD 2111 2111
FT CARBOHYD 2142 2142
FT CARBOHYD 2332 2332
FT CARBOHYD 2533 2533
FT CARBOHYD 2575 2575
FT CARBOHYD 2692 2692
FT CARBOHYD 2812 2812
FT CARBOHYD 3052 3052
FT CARBOHYD 3065 3065
FT CARBOHYD 3144 3144
FT CARBOHYD 3172 3172
PFam; PF02345; TILs; 25.
PFam; PF00094; vwd; 4.
SMART; SM00181; EGF; 2.
SMART; SM00274; FOLN; 11.
SMART; SM00137; MAM; 2.
SMART; SM00214; VNC; 17.
SMART; SM00216; VWD; 4.
PROSITE; PS00022; EGF_1; 1.
PROSITE; PS01186; EGF_2; 18.
PROSITE; PS00026; EGF_3; 1.
PROSITE; PS00740; MAM_1; FALSE_NEG.
PROSITE; PS00060; MAM_2; 3.
Signal; Glycoprotein; Transmembrane; EGF-like domain; Cell adhesion;
Repeat.
SIGNAL 1 17 POTENTIAL.
CHAIN 18 5376 ZONADHESIN.
DOMAIN 18 5310 EXTRACELLULAR (POTENTIAL).
TRANSMEM 5311 5337 POTENTIAL.
DOMAIN 5338 5376 CYTOPLASMIC (POTENTIAL).
MAM 1. 45 210
DOMAIN 215 374
MAM 2. 377 542
DOMAIN 547 1170
DOMAIN 1171 1280
DOMAIN 1281 1669
DOMAIN 1670 2056
DOMAIN 2057 2459
DOMAIN 2460 2579
DOMAIN 2580 2699
DOMAIN 2700 2819
DOMAIN 2820 2939
DOMAIN 2940 3059
DOMAIN 3060 3179
DOMAIN 3180 3299
DOMAIN 3300 3416
DOMAIN 3417 3536
DOMAIN 3537 3656
DOMAIN 3657 3776
DOMAIN 3777 3892
DOMAIN 3893 4028
DOMAIN 4029 4148
DOMAIN 4149 4263
DOMAIN 4264 4383
DOMAIN 4384 4503
DOMAIN 4504 4623
DOMAIN 4624 4743
DOMAIN 4744 4863
DOMAIN 4864 5261
DOMAIN 5259 5295
DISULFID 5263 5274
DISULFID 5268 5283
DISULFID 5285 5294
CARBOHYD 339 339
CARBOHYD 499 499
CARBOHYD 1216 1216
CARBOHYD 1239 1239
CARBOHYD 1314 1314
CARBOHYD 1814 1814
CARBOHYD 1908 1908
CARBOHYD 1933 1933
CARBOHYD 2029 2028
CARBOHYD 2111 2111
CARBOHYD 2142 2142
CARBOHYD 2332 2332
CARBOHYD 2533 2533
CARBOHYD 2575 2575
CARBOHYD 2692 2692
CARBOHYD 2812 2812
CARBOHYD 3052 3052
CARBOHYD 3065 3065
CARBOHYD 3144 3144
CARBOHYD 3172 3172

```

Query Match 4.1%; Score 335; DB 1; Length 5376;

Best Local Similarity 20.3%; Pred. No. 5.1e-14;

Matches 325; Conservative 154; Mismatches 523; Indels 596; Gaps 97;

QY 18 CEHALQPTIEQ---ACLIICPRDVCVSEPLPWSNCSKGGCKLQHETRAVIAPPLPGL 74

DB 2933 CATITLQCPAHSHTNCLPPCQPSCLDSE---GHC-----2964

QY 75 QCPNLTESRADAPISCPGLG---EEY-TFSLKVGPMKCRRLPHLKEI-NPSGRTVLDFN 129

DB 2965 -----EGSITKAPSAQCGCVCEPDVVLNNKCVPAIECGCKDAQGVLPADKTI--N 3016

QY 130 SDSNERTVFKH---QSYKAHHHSWAIEIGYQTRQVSCSTRSDQ---NAMLISLQDSF 183

DB 3017 RGCTQSCITCKGGAIQCFQCPSETYCKDI--EDGNSNCTRIISLQCPANSNFTSCLPSCQ 3074

QY 184 PLTVQSCIT-MFKDCETQWSWSWSPCK--TCRSGLLPGRSRNRNVKMAIGGKCEPE 240

DB 3075 P-----SCSNTDVHCEGSPNTLSSCREGCVQSG-----Y 3105

QY 241 LLEKACIVEGELLQQCPRYSWRTSEWKQVSLLEQDPHMHVTPVC-----GGG 293

DB 3106 VLHNDKCI---LRNQ-----GCKDAQGALIPGKTI--WITSG--CTQSCNCTGGA 3149

QY 294 IQTR-----EVYC-----AQSVPAALRAKEVS-----RPV--EKALCVGPAP 330

DB 3150 IQCNFQCPKATYCKDLKDGSSNCTNIPLOCPAHSRYTNCLPSCPPLCLDPEGLCEGTSP 3209

QY 331 -LPSQLCNIPCTDCIVSSWSWANGLCIHEN-----CHEPQKKGFRTR-----QRH 375

DB 3210 KVPSS---TCREGCIQCP---GYLAKHNKCVLRIFCGCKNTQG--AFISADKTIWSRG 3258

QY 376 VLMESTGPAG--HCPHLVESVPCEDPVCYWLASEGICFPDHGKCGLGHRLKAVCON-D 432

DB 3259 CTQSCITCPAGAIHCRNF-----KCPSG-----TYCKNGD 3287

QY 433 RGEDVSGSL---CP-----VPPPPERKSCETPCRMDCVLSWTWSSCSQSC----- 476

DB 3288 NGSSNCTEITLQCTNTOFTDCLP-----SCVPSCSNRCEVTSVPVSSCREGLCNHG 3341

QY 477 -----SNKSDGKQTRSRITLALAGEGKPPSQALQEHRL--CNDHSCMLHWET 526

DB 3342 FVFSDEKCVPRTOCGCKDARGAIIIPAGKTWTSKGTSCACACVEGNIQQCNFQC----- 3394

QY 527 SPWPCSEDTLVLTALNATIGNGBEATCGVGIQTRRVFCVSHVQVMTKCPDSTRPETV 586

DB 3395 PPTYCKDNS-----EGSSTC-----TKITLQCPAHTQYTSC 3426

QY 587 RPFCLP-----CKKDCI-----VT 600

DB 3427 LPSCILPSCLDPEGLCKDIPKVPSTCKGCVQSGYVLNSDKVLAECCKDAQGALIP 3486

QY 601 AFSEWTP--CPRMCOA--GNATVKOSRYEIIIOEAANGGQECPTLYEERECEDVSL--CPV 656

DB 3487 AGKWTSPGCTQSCACMGAVQCS-----SQPPETICKNEDGNSNCAKITLQCPA 3539

QY 657 YRWKPKQWSPCI---LVPESVWQGITGS--SEACGKGLQTRAVSCISDDNRSAMMECL-- 710

DB 3540 HSLFTNCLPCLPSCLDPDGLCKGASPKVPSTCKEG-----CICQSGYVLNSNKKLLR 3592



```
PT DISULFID 665 677 BY SIMILARITY.
PT DISULFID 672 690 BY SIMILARITY.
PT DISULFID 684 699 BY SIMILARITY.
PT DISULFID 725 737 BY SIMILARITY.
PT DISULFID 732 750 BY SIMILARITY.
PT DISULFID 744 759 BY SIMILARITY.
PT CARBOHYD 88 88 N-LINKED (GLCNAC. . .) (POTENTIAL).
PT CARBOHYD 309 309 N-LINKED (GLCNAC. . .) (POTENTIAL).
PT CARBOHYD 409 409 N-LINKED (GLCNAC. . .) (POTENTIAL).
PT NON TER 867 867
SQ SEQUENCE 867 AA; 91817 MW; 9538F2108E787B49 CRC64;

Query Match 3.98; Score 324; DB 1; Length 867;
Best Local Similarity 21.38; Pred. No. 3.6e-14;
Matches 232; Conservative 90; Mismatches 355; Indels 414; Gaps 58;

2Y 29 QACL-IPCDRCVVSEFLPWSNKGK-GKKLQHRTRAVIAPPLFGGLQC-----PNL 79
DB 19 RACFAAACPDGVTWTSRNSPSEPCGGVTARHE---CHFPQNGRICATILPGGFPST 75
2Y 80 TETRAC-----DAPISCPLEBEYTFSLKVP-----WSKCRPLHLKEI 118
DB 76 RETRPCQDGCNPVTCSEBELVFHACVPCPLTCDDISQATCPDRPCGGPGCWCPCAGQVL 135
119 NPSGRTVLDNSDSNERVTFKHQSYKAHHHSKWAIEIGYQTRQVSCTRSDGO----- 171
DB 136 GAQRCV-----W-----PRQCPL-VDGRYWPQGR 161
2Y 172 ---NAMLSLCLQDSFPPLTVQ---SCIMPXDCETSWSSPSCSKTCRSGLLFGFRSR 225
DB 162 VKTDCQLCVC-QDGRPRRCQPSLDCAV--NCGWSAWSPAECLGPGCSRVQSPSPN- 217
226 NVKHMATGGKCEPPELLEK-----EACIVEGELLQOCPRYSWRTSEWKECQVSLLL 276
DB 218 --NPRPAGRHQCRGLHRKARRCOTECCECEODGRVHRVGER--WRAGPCRVCO---CL 270
277 EQDDPHHVTGPGCGG-----IQRETVYCAQSVPAALRAKEVRPVEKALCVGP 328
DB 271 HDGSAR---CSPYCPGLGSCPDQVNLVEGVGSCCHCVPPGENCTVHPMATPV-----P 320
2Y 329 APIPSQLNCIPCTDCIVSWSAGLCIHNCHPEQKGGFRTRQRHVLMESTGPAGHCP 388
DB 321 APTSPQIGAPLITYLLPPP-----GDPCYSPGLARL-----PEGSIF 359
389 HLVESVCEBPMCMVRLASEGICFPDHBKCGLGHRLKAVQNDREDVSGSLCPVPPP 448
DB 360 --ASSQLEHP---AWAA-----ILR-----PAGAP 381
449 ERKSCETPCRMDCVLSWTWSSCSGSKNS-----DGKQTRSRITLALAGEGKP 501
DB 382 -----GWSPVHADTQGHTPPPVYLQDLQPRNLGIIIVQAGS-- 420
502 CPPSQALQEHRLCNDHSCMLHWETSPWGPCSEDTLVLTALNATIGNGEATCGV---GLO 558
DB 421 ---SDMLQV-----SSDGLHWS---YRDIQGTQAPAPLPKQNWNGPSTVMFARMVQ 468
559 TRRVFCVSHVQVMTKCRDSTRPRTVPCFLPCKKDCIIVTAFSEWTPCPRMCQAGNAT 618
DB 469 ARHVRVWPSD-GHHQAAPSSDANLDGLPLRVLLGCE-----PAPLCIGVG--- 512
619 VKQSRPIIIQEAR-----NGQCECPDTLYEE----- 645
513 -----HRCVSGEAPGAPCDGVDECKDGDGCVTPPAGAGRIESTAWSSAPSAQPG 567
646 ---RECEDEVSLCPVYRWKQKSPCILVPE-----SVMGITGSGSE 683
568 QLPPQPSGELAABEAADWHFGRSP---VPTGKGASLGSSEPHSPGGSV-QTVTPTSQ 623
2Y 684 ACKGL--QTRAVSCISDDNRSAMMECLKQTNMPLLVQECT---VPC----- 727
DB 624 PEAQALAFEMAAVTVLPFHPMVTPEVAGRSTTTPGPPHVVQCSPGQVPCPEVLGCVELEQL 683
728 --REDCTF-----TAWSKFT-----PCSTNCEATKSRRLQLTGSKRKEKQDSLDLYP 773
```

```
DB 684 CDGREDCLDGSDBERPCAWAAGTVFTVPTTT-----LPGLFASRLDLCSPSL-- 730
QY 774 LVEVELCPCEDEFISQPGNWSDCILPEGRREPHRGLRVQADSKCEGGLRFRVACSDKN 833
DB 731 -----TCGSGE-----CLPVERR-----CDLQLD-----CQDG-----SDEN 757
QY 834 GRPVDPSFCSSSGYIOEKVCVPCFPDCKLSDWSWGSWSSCGIGVIRSKWLKEKPYNG 893
DB 758 G-----CV-----DCGLAFWSGWSGSSRSCSLGLGLAFQRELLRPLPLPG 795
QY 894 GRPCPKLDLKNQVHEAVPCYSECNQY--SWVV--RHWSCKNNELRLRCGGGTQSRKIR 950
DB 796 G-SCPPDRLRSQ-----PCFVQACFVAGAMAEWAWGFC-----SVSCGGGHRSRRS 842
QY 951 CVNTADGEGGA 961
DB 843 CNDPPPKNGGA 853

RESULT 14
AT17 HUMAN
ID AT17 HUMAN STANDARD; PRT; 1095 AA.
AC Q8T56;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE ADAMTS-17 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 17) (ADAM-TS 17) (ADAM-TS17).
GN ADAMTS17.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21856482; PubMed=11867212;
RA Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,
RA Lopez-Otin C.
RT "Cloning, expression analysis, and structural characterization of
RT seven novel human ADAMTSs, a family of metalloproteinases with
RT disintegrin and thrombospondin-1 domains."
RT Gene 283:49-62(2002).
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -!- TISSUE SPECIFICITY: Expressed in fetal lung, in adult brain,
CC prostate, and liver.
CC -!- PTM: The precursor is cleaved by a furin endopeptidase (By
CC similarity).
CC -!- SIMILARITY: Belongs to peptidase family M12B.
CC -!- SIMILARITY: Contains 1 disintegrin-like domain.
CC -!- SIMILARITY: Contains 1 PLAC domain.
CC -!- SIMILARITY: Contains 5 TSP type-1 domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ315735; CAC86016.1; -.
CC MEROPS; M12.027; -.
CC
CC Genew; HGNC:17109; ADAMTS17.
CC MIM; 607511; -.
CC InterPro; IPR001762; Disintegrin.
CC InterPro; IPR004818; Pept_M10A_M12B.
CC InterPro; IPR006025; Pept_M_Zn_BS.
CC InterPro; IPR001590; Peptidase M12B.
CC InterPro; IPR002870; Peptidase M12B_N.
CC InterPro; IPR000884; TSP1.
CC InterPro; IPR008085; TSP_1.
```



DR Pfam; PF01562; Pep M12B protease; 1.  
 DR Pfam; PF01421; Reptolysin; 1.  
 DR Pfam; PF00090; tsp.1; 5.  
 DR PRINTS; PR01705; TSP1REPPEAT.  
 DR SMART; SMO0209; TSP1; 5.  
 DR PROSITE; PS00215; ADAM\_MBP; 1.  
 DR PROSITE; PS00546; CYSTEINE SWITCH; FALSE NEG.  
 DR PROSITE; PS00427; DISINTEGRIN 1; FALSE NEG.  
 DR PROSITE; PS00214; DISINTEGRIN 2; FALSE NEG.  
 DR PROSITE; PS00900; PLAC; 1.  
 DR PROSITE; PS00092; TSP1; 5.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;  
 KW Repeat; Extracellular matrix.  
 FT SIGNAL 1 27  
 FT PROPEP 28 223  
 FT CHAIN 224 1095  
 FT DOMAIN 224 452  
 FT DOMAIN 453 542  
 FT DOMAIN 543 598  
 FT DOMAIN 599 701  
 FT DOMAIN 702 779  
 FT DOMAIN 800 860  
 FT DOMAIN 861 922  
 FT DOMAIN 925 968  
 FT DOMAIN 972 1029  
 FT DOMAIN 1045 1084  
 FT DOMAIN 60 120  
 FT SITE 201 201  
 FT METAL 389 389  
 FT ACT SITE 390 390  
 FT METAL 393 393  
 FT METAL 399 399  
 FT CARBOHYD 167 167  
 FT CARBOHYD 483 483  
 FT CARBOHYD 785 785  
 FT CARBOHYD 790 790  
 FT CARBOHYD 832 832  
 FT CARBOHYD 839 839  
 FT CARBOHYD 894 894  
 SQ SEQUENCE 1095 AA; 121099 MW; A56540484754D5D CRC64;

Query Match 3.9%; Score 323; DB 1; Length 1095;  
 Best Local Similarity 21.3%; Pred. No. 5.5e-14;  
 Matches 194; Conservative 66; Mismatches 258; Indels 392; Gaps 51;

QY 85 CDAPISCPGEE-----YTSKVGPNKSLRPHLKEINPSGRVTLDFNSDSNERTFK 139  
 DB 367 CSAKKVCVLAEDNGLNLAFTLAHEL-----HNLGMHDDDD----- 402  
 QY 140 HQSYKAHH-----SKSWALEIGYQTRQVSCSTRSDQN-----AMLSLCLQDSF 183  
 DB 403 HSSCAGRSHIMSGENWKNFSDLSWS-----SCSRDDLENFLKSKVSTCLLVTD 452  
 QY 184 PLTVSCIMPDCETQSWNSPCKTSRSGSLPLGFRSR-SRNVKMAIGGKCEPELL 242  
 DB 453 PRSQHTVPLPHKLPMHYSAEQ-----QILFGMNAFCRNVHELMCAG----- 497  
 QY 243 EKEACIVEGELLQCPRYSWTSEWKEQVSLLEQDDPHHVTGPVCGGIGIOPREYCA 302  
 DB 498 --LWCLVSGD-----TSCKTKL-----DP--PLDGTCEGA-----DKWC- 527  
 QY 303 QSVFAAALRAKESRVPVEKALCVGPAPLPSCNLNIPGSTDCIVSSWSANGLCIHENCHE 362  
 DB 528 -----RAGE-----CVSKTPIPEHV-----DGDWSPWGAWSMC-SRTC-- 559  
 QY 363 PQGKGGRTORHVLWSTGPAG-HCPHL-VESVPCEDPMCYRWLAS----- 407  
 DB 560 ---GTGARFRKCDNPPPGGTHCPASVEHAVENLPCPKGLPSFRDQCOAHRLS 616  
 QY 408 -----EGICFP-----DHGKC----- 418  
 DB 617 PKKGLLTAVVDDKPCLEYSLPKESPLLVADRVLDGTGPGPYETDLCVHGKCRQIGC 676

QY 419 -GLGHRILK-----AVQOND-----RG-----EDVSGSL----- 441  
 DB 677 DGIISAAKEDRCGVCSGDGKTCHLVKGFDSHAGTALKDSKGSINSWKIELPGBEQI 736  
 QY 442 -----CPVPPPPKRSCEIP 456  
 DB 737 ACTTVYVRRGLWEKISAKGPTKPLHLWVLLFFHDQDYGHIHYETVFNRTAENQS--EP 794  
 QY 457 CRMDCVLSWTE--WSSCSQCSNKNKSGKTRSRITILALAGEGK-----PCP-PS 505  
 DB 795 EKPQSLFIWTHSGWEGSVQGG-----GER--RTIVSCTRIVNKITLIVNDSDCPOAS 847  
 QY 506 QALQEHRLCNHSCMLHWETSPMGPCSEDTLVLTALNATIGMNGEATCGVGIQTRRVFCV 565  
 DB 848 RPEQVRRCNLHPC-QSRWAGPWSPCS-----ATCEKGFQHREVTVCV 889  
 QY 566 -----KSHVGQVMTKRCPSDSTRPETVRPCFLPCKKDCI-VTAFSEWTPCPRMCOGNATV 619  
 DB 890 YOLQNGTHA-TRPLYCP-GPRPAVQSC-----GQDCLSIWEASEWSQCSASC--GKGVW 942  
 QY 620 KQSYRIIIQEAANGQCECPDTLYEERECEDEVSLCPVYRWKPKQKWSPCILVPESVWQGIT 679  
 DB 943 K-----RTVACTNSQKCKDASTRPAEAEACEVSGC--YEWKTGDWSTC----- 984  
 QY 680 GSSEACGKGLQTRAVSCISDD-----NRSAEMMECLKQINGMPLLVQECTVP-- 726  
 DB 985 --SSTCGKGLQSRVVQCMHKVTVGRHSGECPALSKPAPYRCYQEVNDRINANTITSRL 1042  
 QY 727 --CREDCFTTAWSKFTPCSTNCEATKSRRLQUTGSRKKEKQPSDLYPLVETELC--PC 782  
 DB 1043 AALTYKCTEDQWTV-----CRVI-----REKNLCQDMRWY-----QRCCQTC 1080  
 QY 783 DEFISQPYGN 792  
 DB 1081 RDF-----YAN 1086

## RESULT 15

AT10\_HUMAN  
 ID AT10\_HUMAN STANDARD; PRT; 1077 AA.  
 AC Q9H324;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE ADAMTS-10 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 10) (ADAM-TS 10) (ADAM-TS10) (Fragment).  
 DE ADAMTS10.  
 GN Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 [1] RN  
 RP SEQUENCE FROM N.A.  
 RA Apte S.S.;  
 RT "ADAM-TS10: a novel member of the ADAM-TS family containing multiple thrombospondin type I repeats."  
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- COPACATOR: Binds 1 zinc ion per subunit (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (By similarity).  
 CC -1- DOMAIN: The spacer domain and the TSP type-1 domains are important for a tight interaction with the extracellular matrix (By similarity).  
 CC -1- SIMILARITY: Belongs to peptidase family M128.  
 CC -1- SIMILARITY: Contains 1 disintegrin-like domain.  
 CC -1- SIMILARITY: Contains 1 PLAC domain.  
 CC -1- SIMILARITY: Contains 5 TSP type-1 domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way



GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

XM protein - protein search, using sw model

Run on: February 24, 2004, 00:30:51 ; Search time 110.308 Seconds  
(without alignments)  
4190.395 Million cell updates/sec

Title: US-10-022-710-2

Perfect score: 8241

Sequence: 1 MVRCTQKLNRTVAVNEICEH.....QSTPPQKPLFLAYDGLDM 1465

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL.25.\*  
1: sp.archaea.\*  
2: sp.bacteria.\*  
3: sp.fungi.\*  
4: sp.human.\*  
5: sp.invertebrate.\*  
6: sp.mammal.\*  
7: sp.mhc.\*  
8: sp.organelle.\*  
9: sp.phage.\*  
10: sp.plant.\*  
11: sp.podent.\*  
12: sp.virus.\*  
13: sp.vertebrate.\*  
14: sp.unclassified.\*  
15: sp.virus.\*  
16: sp.bacteriap.\*  
17: sp.archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8227	99.8	1536	Q9C014	Q9c014 homo sapien
2	4361.5	52.9	1107	11 Q8BHP3	Q8bhp3 mus musculu
3	4296	52.1	1502	4 Q9UPZ6	Q9upz6 homo sapien
4	4108	49.8	740	4 Q8WU00	Q8wuv0 homo sapien
5	1352.5	16.4	446	4 Q43384	Q43384 homo sapien
6	725.5	8.8	4998	11 Q8CG65	Q8cg65 mus musculu
7	720.5	8.7	5146	6 Q8SPM4	Q8spm4 bos taurus
8	712	8.6	3869	5 Q86PQ3	Q86pq3 cryptospori
9	694	8.4	2165	5 Q19791	Q19791 caenorhabdi
10	661	8.0	1637	6 Q9XSU8	Q9xsv8 bos taurus
11	555.5	8.0	4123	4 Q7S851	Q7s851 homo sapien
12	587.5	7.1	1461	5 Q8MTA8	Q8mya8 caenorhabdi
13	555.5	6.7	182	4 Q43376	Q43376 homo sapien
14	474	5.8	1089	5 Q8T3A0	Q8t3a0 ciona intes
15	470	5.7	1121	5 Q966P9	Q966p9 caenorhabdi
16	455.5	5.5	1353	5 Q9V9Q7	Q9v9q7 drosophila

17	449.5	5.5	1148	5	Q966Q0	Q966q0 caenorhabdi
18	442.5	5.4	724	4	Q94862	Q94862 homo sapien
19	440.5	5.3	701	11	Q8CDB8	Q8cdb8 mus musculu
20	440.5	5.3	807	6	Q9GLX9	Q9glx9 bos taurus
21	436	5.3	1043	5	Q90884	Q90884 caenorhabdi
22	434.5	5.3	807	4	Q9HCB6	Q9hcb6 homo sapien
23	427.5	5.2	402	11	Q8K2Q8	Q8k2q8 mus musculu
24	427.5	5.2	807	11	Q8VCC9	Q8vcc9 mus musculu
25	425.5	5.2	807	4	Q8NCD7	Q8ncd7 homo sapien
26	419	5.1	802	13	Q9W770	Q9w770 gallus gall
27	417.5	5.1	808	13	Q42113	Q42113 brachydanio
28	416.5	5.1	622	4	Q9H8X0	Q9h8x0 homo sapien
29	414.5	5.0	803	13	Q42114	Q42114 brachydanio
30	405.5	4.9	1487	5	Q8MPV5	Q8mpv5 caenorhabdi
31	405.5	4.9	1958	5	Q8I710	Q8i710 caenorhabdi
32	405.5	4.9	2167	5	Q76840	Q76840 caenorhabdi
33	405	4.9	1572	5	Q44938	Q44938 haemorchus
34	398.5	4.8	660	5	Q23832	Q23832 cryptospori
35	392	4.8	5636	4	Q96RW7	Q96rw7 homo sapien
36	389	4.7	2772	5	Q9VAV4	Q9vav4 drosophila
37	387	4.7	2673	4	Q96SC3	Q96sc3 homo sapien
38	386	4.7	2174	5	Q9GQR0	Q9gqr0 drosophila
39	385	4.7	2776	5	Q869A0	Q869a0 drosophila
40	385	4.7	2898	5	Q868Z9	Q868z9 drosophila
41	377.5	4.6	821	5	Q22631	Q22631 caenorhabdi
42	373	4.5	1600	11	Q811B3	Q811b3 mus musculu
43	368.5	4.5	687	5	Q23729	Q23729 cryptospori
44	368.5	4.5	1704	5	Q94446	Q94446 chironomus
45	364	4.4	1014	5	Q9V9Q9	Q9v9q9 drosophila

## ALIGNMENTS

### RESULT 1

Q9C014 PRELIMINARY; PRT; 1536 AA.  
 ID Q9C014;  
 AC Q9C014; (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 24, Last annotation update)  
 DE Hypothetical protein KIAA1679 (Fragment).  
 GN KIAA1679.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21082932; PubMed=11214970;  
 RA Nagase T., Kikuno R., Hattori A., Kondo Y., Okumura K., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. XIX.  
 RT The complete sequences of 100 new cDNA clones from brain which code  
 RT for large proteins in vitro."  
 RL DNA Res. 7:347-355(2000).  
 DR EMBL; AB051466; BAE21770.1; -.  
 DR InterPro; IPR000884; TSP1.  
 DR Pfam; PF00090; tsp.1; 8.  
 DR SMART; SM00209; TSP1; 16.  
 DR PROSITE; PS50092; TSP1; 12.  
 FT Hypothetical protein.  
 FT NON TER 1  
 SQ SEQUENCE 1536 AA; 171242 MW; B5121C8160CEDCE CRC64;

Query Match 99.8%; Score 8227; DB 4; Length 1536;  
 Best Local Similarity 95.8%; Pred. No. 0;  
 Matches 1464; Conservative 1; Mismatches 0; Indels 2; Gaps 1;  
 QY 1 MVRCTQKLNRTVAVNEICEHFALQPTPEQACLPDRDGVVSEPLPWSNCSKGGCKLQH 60  
 Db 70 MVRCTQKLNRTVAVNEICEHFALQPTPEQACLPDRDGVVSEPLPWSNCSKGGCKLQH 129  
 QY 61 RTRVIAAPPLFGGLQCPNLTESRACDAPISCLGEEBETFSLVKVGPPWSKCRLLPHLKEINP 120

130 RTRAVIAPLEFGLQCPNLTSSRADAIFSCPLGEEVTFPSLKVGPWSKRLPHLKEINP 189  
121 SGRVTLDNSNSRNVTFKHQSYKAHHHKSWAIEIGYQTRQVSCITSDGQNAWLSLCIQ 180  
190 SGRVTLDNSNSRNVTFKHQSYKAHHHKSWAIEIGYQTRQVSCITSDGQNAWLSLCIQ 249  
181 DSFPITVQSCIMPDCETSOSSWSPCKTKCRSGSLLPGRFSRNRNVKHWAIAGGKECPE 240  
250 DSFPITVQSCIMPDCETSOSSWSPCKTKCRSGSLLPGRFSRNRNVKHWAIAGGKECPE 309  
241 LLEKEACTIVBELLOOCPRYSWRSEWKECOVSLLEQDDPHWHVTGPVCGGIGTREVY 300  
310 LLEKEACTIVBELLOOCPRYSWRSEWKECOVSLLEQDDPHWHVTGPVCGGIGTREVY 369  
301 CAQSVPAALAAKAEVSPVEKALCVGPAPLPSQLCNIPCTDCIVSSWAWGLCIHENC 360  
370 CAQSVPAALAAKAEVSPVEKALCVGPAPLPSQLCNIPCTDCIVSSWAWGLCIHENC 429  
361 HEPOGKGFRTORHVLMESTGPAHCHPLHVESVPCDDPMCYRWLASGICFPDGHKCGL 420  
430 HEPOGKGFRTORHVLMESTGPAHCHPLHVESVPCDDPMCYRWLASGICFPDGHKCGL 489  
421 GHRILKAVCONDRGEDVSGSLCPVPPPERKSCBIPCRMDCVLSEWTEWSSCSQCSKN 480  
490 GHRILKAVCONDRGEDVSGSLCPVPPPERKSCBIPCRMDCVLSEWTEWSSCSQCSKN 549  
481 SDGQTRSRITLALAGEGKCPSPQALQEHRLCNDHSCMOLHWETSPWGPCSEDTLVTA 540  
550 SDGQTRSRITLALAGEGKCPSPQALQEHRLCNDHSCMOLHWETSPWGPCSEDTLVTA 609  
541 LNATIGWGEATCGVIGTQTRVFCVKSHGVGVMTKRCPSDSTRPETVAPCELPCKKCIQVT 600  
610 LNATIGWGEATCGVIGTQTRVFCVKSHGVGVMTKRCPSDSTRPETVAPCELPCKKCIQVT 669  
601 AFSEWTPCPRMCQAGNATVKQSYRIIIQEAANGQECPCDTLYEERCEVDVSLCPVYRWK 660  
670 AFSEWTPCPRMCQAGNATVKQSYRIIIQEAANGQECPCDTLYEERCEVDVSLCPVYRWK 729  
661 POKHSPCLIVPESVWQITGSBAGKGLQTRAVSCISDDNRSAMMECLKQTNWPLLV 720  
730 POKHSPCLIVPESVWQITGSBAGKGLQTRAVSCISDDNRSAMMECLKQTNWPLLV 789  
721 QECTVPCREDCTFTAWSKFTPCSTNCAATKRRRQLTGKSKKEKQCDSDLYPLVETELC 780  
790 QECTVPCREDCTFTAWSKFTPCSTNCAATKRRRQLTGKSKKEKQCDSDLYPLVETELC 849  
781 PCDEFISQPYGNWSDCILPEGRREPHRGLRVQADSKGEGRLRFAVACSDXNGRPVDP 840  
850 PCDEFISQPYGNWSDCILPEGRREPHRGLRVQADSKGEGRLRFAVACSDXNGRPVDP 909  
841 FCSGGYIOEKCVIPCPDPCKLSDWSSWGSSCGIGVIRSKWLKEKPYNGRGPCKL 900  
910 FCSGGYIOEKCVIPCPDPCKLSDWSSWGSSCGIGVIRSKWLKEKPYNGRGPCKL 969  
901 DLKN--QVHEAVPCYSECNOYSWVHWSCKINNELRSALRCGGGTQSRKIRCVNTADGE 958  
970 DLKNOAQVHEAVPCYSECNOYSWVHWSCKINNELRSALRCGGGTQSRKIRCVNTADGE 1029  
959 GGAVDSNLNCODETPPETQSCSLMCPNCEVSEWGLWSKCPQSCDPTMQRTRHLLRPS 1018  
1030 GGAVDSNLNCODETPPETQSCSLMCPNCEVSEWGLWSKCPQSCDPTMQRTRHLLRPS 1089  
1019 LNSRTCAEDSQVQCLLNENCFQYNNLTWSTQCLSENAPCGGVTRLLSCVSDGKP 1078  
1090 LNSRTCAEDSQVQCLLNENCFQYNNLTWSTQCLSENAPCGGVTRLLSCVSDGKP 1149  
1079 VSMDOCEHNLKQPMRISIPCLVECVNQCISGWTAWTECSQTCGHGRMSRTRFIIMPT 1138  
1150 VSMDOCEHNLKQPMRISIPCLVECVNQCISGWTAWTECSQTCGHGRMSRTRFIIMPT 1209  
1139 QEGRPPCPTLTQECTCPVTPCYSWLGNWACKLEGGDCGEGVQIRSLSCWHSIGSISH 1198

1210 QCEGRPCPTLTQECTCPVTPCYSWLGNWACKLEGGDCGEGVQIRSLSCWHSIGSISH 1269  
1199 AGRVEDALCGEMPRODSILKOLCSVPCEGCHLTENSEWSTCELTCTIDGRSFETVGRQS 1258  
1270 AGRVEDALCGEMPRODSILKOLCSVPCEGCHLTENSEWSTCELTCTIDGRSFETVGRQS 1329  
1259 RSRFTIIQSFEENQDSCPOQVLETRPCTGCKVHYTWKASLWNNNERTVWCQRSDGVNVTG 1318  
1330 RSRFTIIQSFEENQDSCPOQVLETRPCTGCKVHYTWKASLWNNNERTVWCQRSDGVNVTG 1389  
1319 GCSPOARPAIQCITPACRKPFSYCTQGVGCGCEKGYTEIMKSNGLDYCMKVPQSEDKK 1378  
1390 GCSPOARPAIQCITPACRKPFSYCTQGVGCGCEKGYTEIMKSNGLDYCMKVPQSEDKK 1449  
1379 ADVKNLSGKNRPNVNSKIHDIFKQWSLQPLDPDGRVKIWWYGVSGGAFIMIFLIFTSLV 1438  
1450 ADVKNLSGKNRPNVNSKIHDIFKQWSLQPLDPDGRVKIWWYGVSGGAFIMIFLIFTSLV 1509  
1439 CKKPPHQSTPPQKPLTLAYDGDLDL 1465  
1510 CKKPPHQSTPPQKPLTLAYDGDLDL 1536

RESULT 2  
Q8BHP3 PRELIMINARY; PRT; 1107 AA.  
ID Q8BHP3  
AC Q8BHP3;  
DT 01-MAR-2003 (TREMBLrel. 23, Created)  
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Weakly similar to hypothetical protein DJ1110N13.1 in chromosome 7.  
DE DI30067103RIK.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxId=10090;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Dorsal root ganglion;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium,  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs."  
RL Nature 420:563-573 (2002).  
DR EMBL; AK051714; SACC34732.1; -  
DR MGD; MGI:2443925; D130067103RIK.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0016491; P:oxidoreductase activity; IEA.  
DR GO; GO:0006118; P:electron transport; IEA.  
DR InterPro; IPR005797; Cytb\_b6\_N.  
DR InterPro; IPR000884; TSP1.  
DR Pfam; PF00090; tsp1; 7.  
DR SMART; SM00209; TSP1; 9.  
DR PROSITE; PS00192; CYTOCHROME\_B\_HEME; 1.  
DR PROSITE; PS00092; TSP1; 10.  
KW Hypothetical protein.  
SQ SEQUENCE 1107 AA; 123640 MW; B34C708B6AA65734 CRC64;  
Query Match 52.9%; Score 4361.5; DB 11; Length 1107;  
Best Local Similarity 83.9%; Pred. NO. 0;  
Matches 759; Conservative 65; Mismatches 80; Indels 1; Gaps 1;  
QY 2 VRCIOKLNRTVAVNICEHFALQPTQACILPCPRDCVWSEFLPWSNCSKCGKGLQHR 61  
DB 143 VRCIOKLNRTVAVNICEHFALQPTQACILPCPRDCVWSEFLPWSNCSKCGKGLQHR 202  
QY 62 TRAVIAPPLFGLQCPNLTESRACDAPISCPGLGEEYTFSLKVGWPKRPLHLKEINPS 121  
DB 203 TRAVIAPPLFGLQCPNLTESRACDAPISCPGLGEEYTFSLKVGWPKRPLHLKEINPS 262  
QY 122 GRTVLDFNSDSNERTVTFKHQSYKAHHHKSWAIEIGYQTRQVSCITSDGQNAWLSLCIQ 181



Db 773 RKRTLVGSKKKKCKNSHLYPLTQVPCDKYNAQVGNWSDCILPEGKVEVLLGMKV 832  
Qy 812 QADSKGEGGRRFRAVACSDKNGRPVDPSPSCSSGYIQEKVCIPCPDCKLSDWSWGC 871  
Db 833 QGDKEGQGVRYAMACYDQNGELVTSRCSNGYIEEACILIPCSDDCKLSEWSWRC 892  
Qy 872 SSSCGIGVIRSKMLKPKYNGRCPCKLDLKN--QVHEAVPCYSECNQYVWVEHSSC 929  
Db 893 SKSCGSGVAVSKVLRKPKYNGRCPCKLDHVNQAQVYVPCVHSDCNQYLVWTEPWSIC 952  
Qy 930 KINNELSLRCGGTQSRKIRCV-NTADGEGGAVDSNLNCQDIPPTQSCSLMCPNECV 988  
Db 953 KVTFFNRENGEGVQIRKVKCMQNTADGSEHVEDYLCPDEENPLGSRCKLPCEDCV 1012  
Qy 999 MSEGLWSKCPQSDPHMTQRTRELLR-PSLNSRTCAEDSQVQPCILLNENCFCQFQYNT 1047  
Db 1013 ISEGWPTQCVLPNCQSGFRQSRADPIRQPADEGRSCPNAVEKEPCNLNKNXYDINVT 1072  
Qy 1048 EWSICOLSENAPCGQVTRLLSCVCSGDKPVSMDQCEQHNLEKPBQMSIPCLVECVNC 1107  
Db 1073 DWSICQISEKAVCGNGIKTRMLDVCVSDGSKVDLKYCEALGLEKNQWQNTSCMVCECPVC 1132  
Qy 1108 QLSGWTAWTECSQTCGHRMSRTRFIIMPTQGEGRPCPTTELTOEKTCPVTPCVSWVLGN 1167  
Db 1133 QLSWSPWSECSQTCGLTGKMRRTVTPQFGDGRPCPSLMDQSKPCVPKCYRQYQG 1192  
Qy 1168 WSACKLEGDCGEGVQIRSLSCMVHSGSISHAAGRVEDALCGEMP-----QDSILKQLC 1222  
Db 1193 WSPQVQAOQCGEGRTRNISCVSDGSDADFSKVPDEECADIELIIDGNKNWLEESC 1252  
Qy 1223 SVPCPGDCHLFEWSEWSTCELTIDGRSFETVGRQSRRTFIIQSFENQSCPOQVLETR 1282  
Db 1253 SQPCGDCYLDKWSWSLQCLTVNGEDLFGGGQVRSRVPVILQELNQHLCPEQMLETK 1312  
Qy 1283 PCTGKCHYHTWASLNMNNERTVWCQRSDGVNTVGGSCPOARPAARQICIPACRKFYSY 1342  
Db 1313 SCYDQCVEYKWMASAMKSGSRVWCORSQDGVNTVGGCLVMSQPDADRSNCPSCOPHSY 1372  
Qy 1343 CTQGGVCGCEKGYTEIMKSNGLDYC----MKVPGSEDKKADVKNLGKNRPNVNSKID 1397  
Db 1373 CSEYKTCHEGGEYEWSSNTLSQCLTILPVVVLPTMEDKGDVKT-SRAVHTQPSNP 1431  
Qy 1398 IFKG--NSLQPLDPGRVKINWYGVSGAGFLIMIFLFTSYLVCKKP-KPHQSTPPQOKP 1454  
Db 1432 AGRGRNFWLPFGDGRKLTWYGVGAAGFLLIFVSMIYLACKPKPKPQRQNNLKP 1491  
Qy 1455 LTLAYDGDLD 1465  
Db 1492 LTLAYDGDADM 1502

## RESULT 4

Q8WUVO PRELIMINARY; PRT; 740 AA.  
AC Q8WUVO;  
DT 01-MAR-2002 (TremBLrel. 20, Created)  
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)  
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)  
DE Hypothetical protein (Fragment).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Muscle;  
RA Strausberg R.;  
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC019344; AAH19344.1; -;  
DR EMBL; BC033125; AAH33125.1; -;  
DR InterPro; IPR000884; TSP1.  
DR Pfam; PF00090; tsp 1; 3.  
DR PROSITE; PS50092; TSP1; 7.

KW Hypothetical protein.  
FT NON\_TER 1  
SQ SEQUENCE 740 AA; 82199 MW; 1E7F2F10FF17794B CRC64;  
Query Match 49.8%; Score 4108; DB 4; Length 740;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 733; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Qy 113 PHLKINPSGRITVLDNFSDSNERTVFKHQSYKAHHHSKSWAIEIGYQTRQVSCSTRSDQN 172  
Db 1 PHLKINPSGRITVLDNFSDSNERTVFKHQSYKAHHHSKSWAIEIGYQTRQVSCSTRSDQN 60  
Qy 173 AMLSICLOSDPPLTVQSCIMPXKDCETSSQSWSPSCSKTCRSGSLLPGRSRNRVVKHAI 232  
Db 61 AMLSICLOSDPPLTVQSCIMPXKDCETSSQSWSPSCSKTCRSGSLLPGRSRNRVVKHAI 120  
Qy 233 GGGKCEPILLEKEACIVGELLQOCPRYSWTSWKECOVSLLEQDDPHHVTGPVCGG 292  
Db 121 GGGKCEPILLEKEACIVGELLQOCPRYSWTSWKECOVSLLEQDDPHHVTGPVCGG 180  
Qy 293 GIOTREVYCAQSVPAALAAKAVSRPVEKALCVGPAPLPQSLNIPGSTDCIVSSWSAW 352  
Db 181 GIOTREVYCAQSVPAALAAKAVSRPVEKALCVGPAPLPQSLNIPGSTDCIVSSWSAW 240  
Qy 353 GLCIHENCHEPGKGFRTORHVLWESTGPAGHCPHLVSVPCEDPMCYRWLASGICF 412  
Db 241 GLCIHENCHDPOGKGFRTORHVLWESTGPAGHCPHLVSVPCEDPMCYRWLASGICF 300  
Qy 413 PDHGKCGLGHRLKAVCQNDRGEDVSGSLCPVPPPPERKSCIEICRMDCVLSEWTESSC 472  
Db 301 PDHGKCGLGHRLKAVCQNDRGEDVSGSLCPVPPPPERKSCIEICRMDCVLSEWTESSC 360  
Qy 473 SOSCKNKSNGKQTSRTILALAGEGKPCPSQALQEHRLCNDHSCMOLHWETSPWGPC 532  
Db 361 SOSCKNKSNGKQTSRTILALAGEGKPCPSQALQEHRLCNDHSCMOLHWETSPWGPC 420  
Qy 533 SEDTLVTALNATIGNGBATCGVGIQTRRVFCVSHVQVMTKRCPCDSTRPETVRPCLP 592  
Db 421 SEDTLVTALNATIGNGBATCGVGIQTRRVFCVSHVQVMTKRCPCDSTRPETVRPCLP 480  
Qy 593 CKKQCVITAFSEWTPCPRMCOAGNATVQSYRIIIOEANGGQCEPDLTYEERECEVDS 652  
Db 481 CKKQCVITAFSEWTPCPRMCOAGNATVQSYRIIIOEANGGQCEPDLTYEERECEVDS 540  
Qy 653 LCPVYRWKPKQKSPCLLVPSVWQGITSSACGKLOTRAVSICISDDNRSAENWECLKQ 712  
Db 541 LCPVYRWKPKQKSPCLLVPSVWQGITSSACGKLOTRAVSICISDDNRSAENWECLKQ 600  
Qy 713 TNGMPELLVQECTVPCREDCTFTAMSKFTPCSTNCEATKSRRLQITGSKRKEKQDSDLY 772  
Db 601 TNGMPELLVQECTVPCREDCTFTAMSKFTPCSTNCEATKSRRLQITGSKRKEKQDSDLY 660  
Qy 773 PLVETELCPDEFISQPYGNWSDCILPEGRPHRGLVQADSKCEGGLFRFRAVACSDK 832  
Db 661 PLVETELCPDEFISQPYGNWSDCILPEGRPHRGLVQADSKCEGGLFRFRAVACSDK 720  
Qy 833 NGRPVPDPSCSSG 846  
Db 721 NGRPVPDPSCSSG 734

## RESULT 5

O43384  
ID O43384 PRELIMINARY; PRT; 446 AA.  
AC O43384;  
DT 01-JUN-1998 (TremBLrel. 06, Created)  
DT 01-JUN-1998 (TremBLrel. 06, Last sequence update)  
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)  
DE Hypothetical protein GS164B05.1 in chromosome 7 (fragment).  
GS164B05.1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.



```
X NCBI_TaxID=9606;
UN [1]
R SEQUENCE FROM N.A.
R TISSUE=Lymphoblast;
R Andrews S., Holmes A., Nguyen C.;
R Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
R EMBL; AC004160; AAC03417.1; -
R InterPro; IPR000884; TSP1.
R Pfam; PF00090; TSP_1; 3.
R SMART; SM00209; TSP1; 7.
R PROSITE; PS00092; TSP1; 3.
R Hypothetical protein.
R NON_TER 1
R NON_TER 446
R SEQUENCE 446 AA; 50444 MW; 0FC13216E84838B9 CRC64;

Query Match 16.4%; Score 1352.5; DB 4; Length 446;
Best Local Similarity 51.0%; Pred. No. 1.2e-119;
Matches 227; Conservative 78; Mismatches 133; Indels 7; Gaps 3;

Y 847 YIOEKVIPCPCDKLSDWSSNGSCSSCGIGVIRIRSKWLKEKPYNGRPPCKLDLKNQV 906
b 1 YIEEACIIPCPCDKLSDWSSNGSRCSKSGGVKVRKWLREKPYNGRPPCKLDLKNQV 60
Y 907 HEAPVCYSCNOYVWYVHWSCKINNELSLRCGGTQSRKICV-NTADGEGGAVDSN 965
b 61 YEVVPCSDCNQLWTEPWSICKVTFVNNRENGEGVQTRKVRQCNQNTADGSPSEHVDY 120
Y 966 LCNQDRIIPETQSCSLMCPNECVSWGLWSKQPCQSDPHTMQRTHLRL-PSLNSRTC 1024
b 121 LCDPEMLPGSRCKLPCPEDCVISEWGPWTQVLPNCQSSFRQSRADPIRQPADEGRSC 180
Y 1025 AEDSQVQPCLLNENCFOFQYNLTENSTCOLSENAPCGQVTRLLSCVSCDKPVMQOC 1084
b 181 PNAVEKEPNLKNKCNHYDYNVDWSTCLSEKAVCGNGIKTRMLDVRSDGSKVDLYKC 240
Y 1085 EOHNEKPMRMSIPICVECVNQLSGWTAWTECSQTCGHGGRMSRTFIIIMFTQEGGRP 1144
b 241 EALGLEKNQMNTSCNVECPVNCQLSDWSPSECSQTCGLTKMIERRVITVQPFQDGRP 300
Y 1145 CPTLTQKTCPTVPCYVSVLGNWSACKLEGGDGGEGVQIRSLSCNVHSGSIHAAGRVE 1204
b 301 CPFLMDQSPCPVKPYRQYQGWSPQEAQCGEGGTRINISCVVSDGSDADFESKVD 360
Y 1205 DALCGEMPE----QDSILKOLCSVPCPGDCHLTWSEWSTCELTICIDGRSFETVGRSR 1259
b 361 EEPFADIELIIDGNKNWVLEESCSQCPGDCYLDKWSWSLCLICVNGEDLGFQGIQVR 420
Y 1260 SRTFIIQSFENQDSCPOQVLETRPC 1284
b 421 SRPVIIQELNQHLCPEQMLETKSC 445
```

RESULT 6  
28CG65

```
ID Q8CG65 PRELIMINARY; PRT; 4998 AA.
AC Q8CG65;
JT 01-MAR-2003 (T-EMBLrel. 23, Created)
JT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
JT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE SCO-Spondin.
EN
SC Mus musculus (Mouse).
SC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
SC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
UN [1]
R SEQUENCE FROM N.A.
R STRAIN=ICR; TISSUE=Brain;
R Goncalves N., Simon-Chazottes D., Creveaux I., Meiniel A.,
R Guenet J.-L., Meiniel R.;
R "Characterization, spatio-temporal expression and chromosomal
R assignment of mouse SCO-spondin.";
```

```
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ491857; CAD42654.1; -
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0005507; F:copper ion binding; IEA.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0005179; F:hormone activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR000923; BlueCu1.
DR InterPro; IPR006207; Cys knot_C.
DR InterPro; IPR008209; EGF_like.
DR InterPro; IPR004421; FAS5_C.
DR InterPro; IPR001545; Gly_HormoneB.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR002919; TIL_Cysrich.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR008085; TSP1.
DR InterPro; IPR005552; VC_out.
DR InterPro; IPR001007; VWF_C.
DR InterPro; IPR001846; VWF_D.
DR Pfam; PF00754; F5_F8_type_C; 1.
DR Pfam; PF00057; ldl_recept_a; 10.
DR Pfam; PF01826; TIL; 10.
DR Pfam; PF00090; TSP_1; 25.
DR Pfam; PF00093; wvc; 1.
DR Pfam; PF00094; wvd; 3.
DR PRINTS; PR00261; LDLRECEPTOR.
DR PRINTS; PR01705; TSP1REPEAT.
DR SMART; SM00192; LDLA; 10.
DR SMART; SM00209; TSP1; 25.
DR SMART; SM00214; VMC; 4.
DR SMART; SM00215; VMC_out; 9.
DR SMART; SM00216; VMD; 2.
DR PROSITE; PS00196; COPPER_BLUE; 1.
DR PROSITE; PS01225; CTCK_2; 1.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS00022; FAS5C_3; 1.
DR PROSITE; PS00261; GLYO_HORMONE_BETA_1; 2.
DR PROSITE; PS01203; LDLRA_1; 8.
DR PROSITE; PS00068; LDLRA_2; 10.
DR PROSITE; PS00092; TSP1; 25.
DR PROSITE; PS01208; VWF_C; 1.
DR PROSITE; PS01084; VWF_C; 2.
SQ SEQUENCE 4998 AA; 535019 MW; DA2ABA8DA47DF225 CRC64;

Query Match 8.8%; Score 725.5; DB 11; Length 4998;
Best Local Similarity 22.3%; Pred. No. 1.4e-58;
Matches 377; Conservative 137; Mismatches 478; Indels 695; Gaps 103;

QY 18 CEHFALQPPTEQAQLIPCRDVCVVSEPLPWSN---CSKGGKQLQHRTRAVIAP-PLFGG 73
Db 3224 CNHCTMGEGLNCTDLFCQ---VSGDKCPWSKWTACSQPCRQTRTSRACVCPAQHGG 3280
QY 74 LQCPNLT-----ESRACDAPISCPLEEYFTSLKVGWPKC-----RLPHLK 116
Db 3281 SPCEESGGTGVQHQEACFNATACPDVGAWSPW---GPWSSCDACLGQSYRSVCSHP 3336
QY 117 EINFSGRTVLDFNSDSNERTVFKHQSYKAHHKSWAIEIGYOTROVSCFRSDQNAMLS 176
Db 3337 PISDGGKPCIG-----GYQQR-PCRNSS-----T 3360
QY 177 LCL-----QDSFPLTVQSCIMP--KDC-ETTSQSSWSPCKTCRSGSLLP----- 218
Db 3361 LCTDCGGGQDLLP-----CQPCRHSQDLISLSTCQPGSAGCGGCGPFGQLSQHGLC 3415
QY 219 -----GFRSSRNV---KMAIGGKCEPILLEK-----EACIVEGELLQ----- 255
Db 3416 VFPVVDCHCFQPRAMGIPENRSRVSGTSLSSWSLEPGEVVTGPCDNCTCAVILQCHEV 3475
QY 256 -QCPR-----YSWRTSEWKECCQVSLLEQQDPHHVHTGPGVCGGGIQTREVVYCAQSVFAAA 310
Db 3476 PSCFGPGIWSSW--GPWEKCSVS-----CGGGEQ----- 3502
```

311 LRAKEVERPEKALCVGAPALPSOLNI-----PCS-----TDCIVSSWSANGLCIH-- 357  
3503 LRSQCARPP-----CPGLAQ-QSRICHIVCRETGCPAGRLYRECQPSDGCPS-CAHVT 3556  
358 -----ENCHEPOGKGFETROHVLWESTGPAHCPLHVESVPC-----ED 398  
3557 QOVACFERCKGCHCFEG-----TFQHVA-----CVQBCPCVLTULLQEL 3599  
399 PMCYRWLASGICPPDHGK-CGLGHRILKA-VQNDRG--EDVSGSLCPVPPPERKSC 454  
3600 GLASAAALGYPTLLGDEKPLGPGVELLPQMLQTDGNCSCVHGKL-----SCS 3649  
455 -IPC-BNDVLSWTEWSSCSNKNKSGKQSRST--ILALAGEGKPCPPSQALQ 509  
3650 MVEGSRVHSGFGPMWLSLCSRCG-----GLGTRTRQCVLPTLAPGGLSC-RGPLJQ 3702  
510 EHLRNDHSCMLHWET-----SPWPGCEDTLVTALNATGMNGE----- 550  
3703 DLEYCFSPGCTAGTIVPTVGLAGKGPWSPSCSHSCTDPAHPA---WRSRTRLC 3759  
551 ATCGVGTQTRVFCVKSHVQVMTKCPDSTRPTVPC-----FLPC-----KKDC 597  
3760 ANCTVG-----DSQEQ--RPNLPSCAALLPCFGPGCGSNC 3794  
598 IVTAFSEWTPCPRMCOAGNATVKQSRVRIIIOEAANGQECPCD--TLYEERECEDVSLCP 655  
3795 FWTSWAPNEPCSRSGVG---QQRRLR-AHPPPGGHWCPDILTAYQERRFCNLRACP 3849  
556 V-----YRWKPKWSPFCLVPSVWOGITGSEACGKGLQTRAVSCISDDNRSAMMECLK 711  
3850 VPGGWSHWP--WSWC-----DRSCGGGRSLRSRSCSPFPKNGG--TSCVG 3892  
712 QTNWMLLVQECT-VPCREDCTFTAMSKFTPCNTCEATKSRRLTGKSRKEKQCDSD 770  
3893 ERHH-----VRCNPMPCBEGC--PAGMEWUSCANHCPCYSCLDQZ-GGMQEQOACOLG- 3944  
771 LYPLVETELCPDE-FISQPYGNWSDCILPEGREPHRGLRVQADSKEGEGRLFRVAC 829  
3945 -----CRCSEGFLEQDGG---CV-PVGHCE-----C 3966  
830 SDKNGRPVDPFCSSSGVIOEKC-----VIPC-----PFDCKLSDWSNGSCSS 874  
3967 TDAQR-----SWAPGSOH-QDACCNNCSQAGQLCTAQLCSPPAHCAWSHSAWSSCSHS 4021  
875 C---GIGVIRIS---KWLKE--KPYNGRCPKLDLKNQVHEAVPC----- 912  
4022 CGPQGOQSRFRSSTSWALEQEQSOSQCPCE-----VPCFPLCLHEHLHEL 4071  
913 -----YSECNOYS-----WVW-EHWSSCKINNELRLSRCCGGTQSRK 948  
4072 GDNWLHGECQCCSTPEGAICKDTDCAVPRGWTLWSSWSYC-----SVSCGGGSQVRT 4124  
949 IRCVNTADGEG----- 960  
4125 RSCVTSAPPHGSLSCGPDQTRHCGQQLCLQKLERCSWGPWGPCSRSCGTGLASRSGC 4184  
961 ----AVDNLNQDEIIPETQSC-SLMCPNFCVSWGLMSKCPQSCDPTHMQRTRHLL 1015  
4185 PCLLTKEKSKNDTFLGLDTQACYSGPCDDCTGDMSSWTRC--SKVLVQCYRHRQVP 4242  
1016 RPSL--NSRTCAE-DSOVQPCLLNENC-----FOFYNLTEWSTCQLSENAFCQGV 1064  
4243 APQAGEGTPTRLDCHFRPTIG-NCSEDSCTPPFFEQ-----SCGSPCAGLC 4290  
1065 RTRLLSCVCSB-----GKPVSMDOQE-----QHNLEKPFQMSI 1097  
4291 ATHNLNRLQDLPQCPGCPKGLLEQAGSCILPEQNCNWHISGEGARVTLAPGDRQL 4350  
1098 PCLVECVV-----NCC-----LSGWTATWC-----SCTC--CH----- 1124  
4351 GC-KEVCRRBELQCSGCGGGLPLTQSWSEWSPGCPCLPOSALAPASRTALEGHPLNT 4409  
1125 -----GGRMSRTRFI--MPTQEGRPCTELTQKTCF-----VTPCYSW 1163

4410 SLDPPPSVTLASEQYRHLCLDPETRPWAGDPALCTVPLSQRLCPDPGACNDTCQWG 4469  
1164 VLGNWSACKLEGDCGEGVQIRSLSCMVHSGSISHAAGRV-EDALCGEMPFDLSILKQLC 1222  
4470 PWPMSPCQM--PCSGGFKLR-----WRVARDTSAGE----- 4499  
1223 SVPCGDCHLTWSEWSTCELTCTIDGRSFETVGRQSRRTFIISFENQDSCPOQVLETR 1282  
4500 ---CPG-----PWAQTESNMGSCFGESECT-----RDTVF---TLDCAQCP-----R 4537  
1283 PCTGGKCYHYTWKASLWNNERTVWCQSRSDGVNVTGGCSPQARPAARQCIPACKKPSY 1342  
4538 SC-----ADLW-----DGVQCLQG-----PCSPGCRCPGQ 4563  
1343 CTQGGVC 1349  
4564 LVQDGHG 4570

RESULT 7  
Q8SPM4  
ID Q8SPM4 PRELIMINARY; PRI: 5146 AA.  
AC Q8SPM4;  
DT 01-JUN-2002 (TREMELrel. 21, Created)  
DT 01-JUN-2002 (TREMELrel. 21, Last sequence update)  
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
DE SCO-spondin.  
GN SCO-SPONDIN.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-Subcommissural organ;  
RA Meinzel A.;  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE-Subcommissural organ;  
RX MEDLINE=20465125; PubMed=11008217;  
RA Gobron S.;  
RT "Subcommissural organ/Reissner's fiber complex: characterization of  
SCO-spondin, a glycoprotein with potent activity on neurite  
outgrowth";  
RL Glia 32:177-191 (2000).  
DR EMBL; A416457; CAC94914.1; .  
DR GO; GO:0005576; C:extracellular; IEA.  
DR GO; GO:0005179; F:hormone activity; IEA.  
DR GO; GO:0007155; P:cell adhesion; IEA.  
DR InterPro; IPR001064; Crystallin.  
DR InterPro; IPR006208; Cys knot.  
DR InterPro; IPR006307; Cys\_knot\_C.  
DR InterPro; IPR000421; FAS5\_C.  
DR InterPro; IPR001545; Gly\_hormoneB.  
DR InterPro; IPR002172; LDL\_receptor\_A.  
DR InterPro; IPR009041; PMP\_inhibitor.  
DR InterPro; IPR002919; TIL\_Cysrich.  
DR InterPro; IPR000884; TSPI.  
DR InterPro; IPR001007; VWF\_C.  
DR InterPro; IPR001846; VWF\_D.  
DR Pfam; PF00007; Cys\_knot; 1.  
DR Pfam; PF00754; F5\_F8\_type\_C; 1.  
DR Pfam; PF00057; ldl\_recept\_a; 10.  
DR Pfam; PF01826; TIL; 10.  
DR Pfam; PF00090; tsp\_1; 25.  
DR Pfam; PF00093; vwc; 1.  
DR Pfam; PF00094; vwd; 3.  
DR PRINTS; PRO0261; LDLRECEPTOR.  
DR SMART; SM00041; CT; 1.  
DR SMART; SM00231; FA58C; 1.

```

JR SMART; SM00068; GHB; 1.
JR SMART; SM00192; LDLA; 10.
JR SMART; SM00209; TSP1; 25.
JR SMART; SM00216; VWD; 3.
JR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.
JR PROSITE; PS01235; CTCK_2; 1.
JR PROSITE; PS01285; FAS8C; 1.
JR PROSITE; PS01286; FAS8C; 1.
JR PROSITE; PS00022; FAS8C; 1.
JR PROSITE; PS00261; GLYCO_HORMONE_BETA_1; 2.
JR PROSITE; PS01209; LDLRA_1; 7.
JR PROSITE; PS00068; LDLRA_2; 9.
JR PROSITE; PS00092; TSP1; 25.
JR PROSITE; PS01208; VWFC; 1.
JR PROSITE; PS00184; VWFC; 2.
JR SEQUENCE 5146 AA; 543576 MW; 724C5FB8727E13DA CRC64;

Query Match      8.7%; Score 720.5; DB 6; Length 5146;
Best Local Similarity 21.3%; Pred. No. 4.3e-58;
Matches 423; Conservative 155; Mismatches 571; Indels 835; Gaps 115;

2Y 28 EQACLIP-CPRDVCVBEFLPMSNCSKGCKKQHRTRAVIAPLFG----- 72
Db 3221 EEPCLLPEDRAGCWGPMWSSCSRSRACDQPPQGLGDCYCEGPRAQGA 3280
2Y 73 --GLOCP--NLT-----ESRACDAPISCLPGEETFSL-KVGNKSKRLPHLKINPSG 122
Db 3281 CQALPCPVNTCTAIEGAIVSACGPP--CPRSCDDLVCHVHCQPGCYC-----PPG 3329
2Y 123 RTVLDNPSNERNVTFKHOSY-----KAHHSKSWAIEIGYQTRVSCSTRSDGNAMLS 176
Db 3330 QVL---SADGTVHVQGHSCDLLLGERHPGAQLAKPDGNC---YCTCSEGGLTCTD 3382
2Y 177 LCLQDSFFLTVQSCINPKDCTSCNWSNPSCKTCHRSGLLPGRFSRSNV-----KMA 231
Db 3383 L-----PCVPWAGWP-----WSEWATACQPCQGT-----RTRSRACAPAQH-- 3422
2Y 232 IGGKECP-----ELLKEACIVSGELLQOCP-RYSWRT-SEWKECOVSL--LLEOOD 280
Db 3423 --GGAPCPGEAGEAGHQRETCASTPE---CPVDGANSWPKSPSPCEVCLGRSHREE 3476
2Y 281 PHHVTG---PVCGGGIQTR-----VYCAQSVPAALAAKAEVSRP 319
Db 3477 CSWPPF--SEGRCPCFGHROSRCPOGNSQCTDCCAGQDILLPCQPCPRS---CEDLSPG 3532
2Y 320 VE-----KALCVGPAPL----- 331
Db 3533 VEQPDSMGCCQPRCCPGQLSQDGLCTPSCRCQYQPGAMGFPENQSRAGSGLSW 3592
2Y 332 ----PSQLCNIPC-STDCI-----VSSWSAWG-----L 354
Db 3593 ESLEPGEVVTGPDNCTCVAGILQCOEVPACSGGLGWSGWPWEDCSVSCGGEGQLRFR 3652
2Y 355 CIHENCHEPGKGFTRQHVLMESTGAGH-----CPHLVESVPCEDPM 400
Db 3653 CRPPCPGPARQS--RTCTQVCREAGCAGRLYRECQSEGPCFSAHVTQGVCFASG 3710
2Y 401 CYRWLASGICFPDGHGKGLGHRLKXAVQND-----RGEDVSGSLC--- 442
Db 3711 C-----EEGCHCEP-----GTLFHSACVQECPCVLTALWQLGGAAGADPGAHLSVLG 3759
2Y 443 ----PVPDPPE-----RKSC-----ETPCRMDCV-----LSEWTESSCSQCS 477
Db 3760 ENGQPLPGDELGSGSLATGCHNCSCAHGLKSCSVSEACSKAAGFSWPGWGPCSRSG 3819
2Y 478 NKNSDGQTRSR--TILALAGEGKPC-PPSQALQEHRLCNDHSCQLHWET----- 526
Db 3820 GL---GTRRESQCVPMAPAGQCGCHGPHWDL---YCPSEPCGAAGSTAEPAITLPG 3873
2Y 527 ----SPWGPCSEDTLVTALNATIGNGEATCGVGTQTRRFVCVSHGVQVMVTKCPDST 581
Db 3874 GWLWSPWSPSCSGTCTDPAHPA---W-----RSRRLCLANCTGGAASQE----- 3915

```

```

QY 582 RPETVRECFILP-----CKK-DCIIVTAFSEWTPCPRMCOAGNATVKQSRVRII 628
Db 3916 ----RPNLPSCCTELPLCPGPGCEAGNCSTWATAWPWPCSGSCGVG-----QQRLLR-AY 3965
QY 629 QEAANGQEGCPD--TLYEERECEDVSLCPV-----YRWKPKM----- 664
Db 3966 HPPGPGGHWCVPDLVTAIQRERFNCNLRACVPVGGWSRSPMSWCDRSCGGRSLRSCSS 4025
QY 665 -----SPCI-----LVPSVMGIGTSGSEACGKGLQTRAVSCIS----- 698
Db 4026 PPPKNGGAPCVGERHARLCPN-----TPCEGCPAGME--VWSCANRPRRCSDLQE 4076
QY 699 ----DNRSANMEC-----LKQING-MPLLVOECT----- 724
Db 4077 GIVQEDQACQCCRCPEGSLEQDGGCVPLGHCECTDAQGHSNAPGSHQHEACNNCTCRA 4136
QY 725 -----VPC--REDCTFTAWSKFTPCSTNCEATKSRRLQLTGKSRKKEKQSDLYPLV 775
Db 4137 QQLSCTAQCPPPPACAWSRSAWSFCSRSCGPAQQSR----- 4175
QY 776 ETELCPDDEFISQPYGNWSDCILLPEGRREPHRGLRVQADSKEG-----EGLRFRA 826
Db 4176 -----FRSSTSGSWA-----PECREE-----QSQPCPQSPCPPLCLQGTFRS 4215
QY 827 VA-----CSDKNRGPVDPSCSSSGVIOEKVIPCFFDCKLSDWSSWSSSCSSCG 876
Db 4216 LGDSWLODGCQCS-----CTPEGLICEDA--ECAGLGAWTFWSPWSDCPVSCG 4262
QY 877 IGVIRSK-WLKEXPYNGRCPKLDLKNQVHBAVPCYSECNQYSVWVHWSCKINNEL 935
Db 4263 GGNQVTRVCVASAPPRGSGFCLGPDVQSQRCLWPCPALPDTCSW--GPWGPC----- 4314
QY 936 RSLRCGGTOSRKIRC-VNTADGEGGAVDSNLNQODEI PPETQSC-SLMCPNCVWSEWG 993
Db 4315 -SRSCGGLASRSASCPCLLAEP-----CNSTSPRLDTQACVAGCPLEECVWSWS 4367
QY 994 LWSKPOSCDPHTMQRRTHLLRPLNSR---TCAB-DSQVQPCLLNENC-----F 1040
Db 4368 SWTRC--SCEV-LVQRYRH-QRPAPGAGAGPCTRLDGHFRPCLTG-NCSEDSACPPF 4422
QY 1041 QFO-----YNLTWSTQLSNAPCGQVTRLLSCVCS-----GKPVMSDQ 1083
Db 4423 EFOAGSFPCTGLCATYLSPW-LCO--DLPECQPG-----CYCEGLLEQAGCVPPEQ 4472
QY 1084 CE-QH-----NLEKPRMSIPCLVECVNQ-----LSGWTAWTEC 1118
Db 4473 CNCHVSGEGAVTLAPGDRLOLGC-KEC--EQRGELOCTSGCCQGLLPLSGWSESPC 4529
QY 1119 -----SOTCG-----HGRMSRTRFIMPTQGE---GRP- 1144
Db 4530 GPCPLPLGLLAPASRAALERWPQDTAGLSPTSAPTLASQHRRLCLDPTGRPWAGDPD 4589
QY 1145 -CPTELTOEKTCPVTP-----CVSWVLGNWSACKL-----EGGDC-GE 1180
Db 4590 LCTVPLSQQLRCP--DPGACQDLQWGPWAGWSCQVPCSGGFLRWREAGIPFGGCRGP 4648
QY 1181 GVQIRLSMCMVHSGSISHAAGRVEDALC-----GE 1210
Db 4649 WAQTESCNMGPCGESCEAODTPTPDCAQCPRSCVDLWDRVECLQGPCRCPCPPGQ 4708
QY 1211 MPFQDS----ILKQLCSVPCG-----DCH-----LTEW 1235
Db 4709 L-VQDGHCVFSSCRCGLPSPFNASWALAPAEVVRLDCRNCTVNGSLACSHECPTLGPW 4767
QY 1236 SEMWTCBLTICDGRSFETVGRQSR-----RTFIQSPENQDSC-----POOVL 1279
Db 4768 SAMENCAPCGGTT-----KBRHSCKEGFGVTPCQAQDMEQQQDCNLQPCPCPPQVL 4822
QY 1280 ET-----RPCT-----GKCC-----VHYTWKAS 1297
Db 4823 SACAVSFCRLCSHLQPGTFCWQBPQCLGCDPCRGQLHNGTCVPPAECPCTQLSLWGLT 4882
QY 1298 LWNNER-----TWMCQRSDGVNVTGG-----CSP-----QARPAAIRQCIPA 1335

```

Db 4883 LTLEEQHRELPPTLLTQNTCHICOGAIFCSLTDCQECPPGETWQVAPGELGCEQT 4942  
2Y 1336 CRKPFSTCTGGV-----CGCEKGYTEMKNGFLDYCMKV-----PGSE 1375  
Db 4943 CREPNATETGNCGRQAPGCVQQRGHRFSGEP-----CVPDLCCECWHGHRPHEPGSE 4997  
2Y 1376 DKKA 1379  
Db 4998 WQKA 5001  
RESULT 8  
Q86PQ3  
ID Q86PQ3 PRELIMINARY; PRT; 3869 AA.  
AC Q86PQ3; 2003 (TREMBlrel. 24, Created)  
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE TSP1 domain-containing protein TSP2 precursor.  
OS Cryptosporidium parvum  
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;  
OC Cryptosporidiidae; Cryptosporidium.  
OX NCBI\_TaxID=5807;  
[1]  
SEQUENCE FROM N.A.  
MEDLINE=22326344; PubMed=12438378;  
Deng M., Templeton T.J., London N.R., Bauer C., Schroeder A.A.,  
Abrahamson M.S.;  
"Cryptosporidium parvum genes containing thrombospondin type 1  
domains";  
RL Infect. Immun. 70:6987-6995 (2002).  
DR EMBL; AY190984; AA039046.1; -  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0005489; F:electron transporter activity; IEA.  
DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
DR GO; GO:0030154; P:cell differentiation; IEA.  
DR GO; GO:0006118; P:electron transport; IEA.  
DR InterPro; IPR001450; 4Fe4S ferredoxin.  
DR InterPro; IPR002086; Aldehyde dehydr.  
DR InterPro; IPR000800; Notch\_dom.  
DR InterPro; IPR003582; Shkt\_dom.  
DR InterPro; IPR00436; Sushi\_SCR\_CCP.  
DR InterPro; IPR000884; TSP1.  
DR Pfam; PF00084; sushi; 1.  
DR Pfam; PF0090; tsp.1; 10.  
DR SMART; SMO0004; NL; 6.  
DR SMART; SMO0254; Shkt; 5.  
DR SMART; SMO0209; TSP1; 11.  
DR PROSITE; PS00198; 4FE4S\_FERREDOXIN; 3.  
DR PROSITE; PS00070; ALDEHYDE\_DEHYDR\_CYS; 1.  
DR PROSITE; PS50092; TSP1; 10.  
KW Signal.  
FT SIGNAL.  
SQ SEQUENCE 3869 AA; 429510 MW; 1P4851B7B0787B87 CRC64;  
Query Match 8.6%; Score 712; DB 5; Length 3869;  
Best Local Similarity 22.8%; Pred. No. 1.9e-57;  
Matches 346; Conservative 153; Mismatches 448; Indels 570; Gaps 81;  
Qy 33 IPCPRDCVSEFFPWSNCSKGCCKLQHRTRAVIAPPLFGL-QCPNLTESRAC---DAP 88  
Db 187 IPCPLSCTVSENGWNSRCSLTGIGHQWRERSVIKAPKDQNLFCQCPETRIQICIDTCS 246  
Qy 89 ISCPGLGEETFTSLKVPWSKCLPLHLKINPGRVLDPNSDNSNRVTFKQSYXAHH 148  
Db 247 SNCTLGFKFKSAVSGSP---CLM-----EKGCVERREILHPFPF--- 284  
Qy 149 SKSWAIGVQTRVCSRTSDGQNALSLCLQDSFFLTQSCIMPKDCETSONSWSPCS 208  
Db 285 -----ELGNFTCKVERHSD-----CLGLHL-----GCAGP---CFEFSWGSLSLCS 323  
Qy 209 KTCASGSLLPGF-----RSRNRVKNHMAIGGKCEPPELLEKEACIVEG 251

Db 324 YYKNEGL---GFDDEASKYLLKKRRVARRMRDA-----NCEDELEYKECDESE 371  
Qy 252 ---ELLQOCPRYSWRTSSEWKECQVSLLLQQDPHWHVTGVPVCGGIGIOTREVYCAQSVPA 307  
Db 372 RHGVREALVRNMGDM---SSWSSCSTS-----CGLSRLRARLLSEFEN 413  
Qy 308 AALRAKEVSRPVEKALCVGPAPLPQCLN-IPGSTD-CIVSSWSAWGLCIHENCHPEQG 365  
Db 414 EEALSENLIN---SESYC-GPL-FQTKACNDRSLTEGCKVSDMGOWSACSCKSNFP-- 466  
Qy 366 KKGFRTRORHVLMESTGAGHCHPLVESVPCEDPMCYRWLASSEGICFPD--GKCG 419  
Db 467 --GLSKQRSVI--SLPRKGDGCPVLESRD-----LGAC---DHSVSVSKNCM 509  
Qy 420 LGHRILKAVQND---RGEDVSGSLCPVPPPPPP-----KS 452  
Db 510 FGDWSSWSTCQDDCYTRYNSSGANRSEKPKRYRHKMIIFNPKDKCKSENYLEVEN 569  
Qy 453 CEIICBMD---CYLSEWTESSQSCSNKNSDKQTRSRITILALAGEGKP----- 501  
Db 570 CEGGCENANVEICQVSEWGPWSPSCSANCDSRRIRIRINSRGRGVKAGSRPLTRGVSS 629  
Qy 502 -CPPSQALQEHRLCNHSCMLHWETSPW-----GPCSEDTLVTALNATIG 546  
Db 630 SCP---SLLEVEKCNTHPC-EYSCELSPWYSNSKSTGPQDRRGNRSDAMLEDERNEDIV 685  
Qy 547 WNGEAT-CGVGI--QTRRV-----FCVZSH--VGQVTKRCPDSTRPTVPCPLCKKDC 597  
Db 686 ISDCSTKCGIGGVERSRIILSGFVSSNEEQNYTSKVG-----FLK--- 729  
Qy 598 IVTAFSEWTPCPRCMQAGNATVKQSRVRIIIQEAANGQECPCDTLYEERECEDVSL-CPV 656  
Db 730 -----DTIKCVRISEGCSDVQCV 747  
Qy 657 YRWPKQKSPILVPESVMQGITSSACGKGLQTRAVSCISDDNRSAEMECUKQTNMG 716  
Db 748 GEM--SSWSAC-----SKSCGEGFQ-----NRSRELL-----V 773  
Qy 717 PLLVQECTV-----PCREDCTFTAWSKFTPCSTNCEATKSRRLQTLGSKRKEK 765  
Db 774 PSLGRKCELAATEELRECFERPCPSSCQVSPSEWTMC----- 810  
Qy 766 QDSDLVPLVETELCPDCEFTSQPYGNWSCILPEGRREPHRGLRVOADSKGEGELRFR 825  
Db 811 -----LGGCDE---KPY-----KRRERKILEAPAPQNC----- 836  
Qy 826 AVACSDKNGRPVDPSPFCSSSGYIOEC--VIPCPDCKLSDWSSWGSCSSCGIGVIRS 883  
Db 837 -LALEEK-----VD-----AEECPRIERCPDCKVGEWSWSECDACKGIGVEKRL 882  
Qy 884 KWLKEKPYNGGRPCPKL-DLKNQVHEAVPCYSECNQYSWVVEHWSCKINNELRSLRCGG 942  
Db 883 RKVVKRESKGGAPCPNLEDLRPCSREA--CKSDCVLGWT--EWGVC-----SKSCGG 931  
Qy 943 GTQSRKTRCVNTADGEGAVDSNLNQDEIIPPETQSCS---LMCPNECVMSWGLASKCP 999  
Db 932 GARSR-VREVISQNEG-----KECEFLKEPEPCNEFCQIATRDCEVQGSWSPCS 982  
Qy 1000 QSCDPHTMQRTRHLRLPSSLNRTCAEDSQVQCLLNENCFQOYNLTENSTCOLSENAP 1059  
Db 983 ASCG-GGVKREQREVQVATGGGR-----EFELN----- 1011  
Qy 1060 CGQVTRKLLSCVSDSKPVSMDQCEQHNLEKFORMS-IPC-LVECVNQLSQTWATE 1117  
Db 1012 --QKVGCGNFKC--PGEP-----C-----IDRPAQEVWPCSIKAMFGCQ----- 1048  
Qy 1118 CSQTCGHGGRMSRTRFTIIMPTQGEGRPCPTTELTEKT---CPVTPCYSVVLGNWSACKL 1173  
Db 1049 -----KRLIDVAKSNGVYPDDRPPEARIMDCCPAT-CGMCV----- 1084  
Qy 1174 EGGDCGEGVQIRLSJSCMVHSGSISHAAGRVEDALCGBMPFQDSILKOLCSVPFGDCHLT 1233

```

1085 ---EAPGQQLRDVGNLSDPACNNAACRFDGDC---KNSPOKASCLLPKLEAF- 1136
1234 EMEWSTCBLCIDGRSFETVGRQSR-----RTIIOSFENQDSCPOQV 1278
1137 -FLSKKEDQVRKNGSDFLEIGGNSGKKEAQNILKSSDNNLVLESFTSQSLNS-- 1193
1279 LETRPTGCKGVHYTWKASLN-----NNEETVACQSDGVNVTGGCSPQAR 1325
1194 -----SSGKSNNGYFAGNQAQVQVHEMFSEQEDYNNR-----HNGDGSN-- 1233
1326 PAAIROCI PACRPFYCTQGGY-CGCEKG-----YTEI-----WKNGFLDYCMKVP 1373
1234 --TIRE-----YKGGDTVIYRCRDKGRKYTAIPYKYCKESGFIQLI----- 1276
1374 SEDKADVKNLSGRNP 1390
1277 -BFGAPLLYVDGDP 1292

RESULT 9
19791
ID Q19791 PRELIMINARY; PRT; 2165 AA.
AC Q19791; Q27524;
CT 01-NOV-1996 (TrEMBLrel. 01, Created)
JT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
OT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE F25H8.3 protein.
IN F25H8.3
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
CX Rhabditidae; Peloderinae; Caenorhabditis.
XX NCBI_TaxID=6239;
[1]
SEQUENCE FROM N.A.
CA Gajadasy S.
IL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
IN [2]

SEQUENCE FROM N.A.
MEDLINE=94150718; PubMed=7906398;
A Wilson R., Almscough R., Anderson K., Baynes C., Berks M., Coulson A.,
A Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fulton L.,
A Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
A Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
A Jones M., Kershaw J., Kirsten J., Laister N., Matreille P.,
A Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
A Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
A Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
A Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
A Watson A., Weinstock L., Wilkinson-Sproat J., Woldman P.,
A "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
A elegans."
A Nature 368:32-38(1994).
IL EMBL; Z69361; CAA93288.1;
A EMBL; Z69360; CAA93288.1; JOINED.
A EMBL; Z69360; CAA93287.1;
A EMBL; Z69361; CAA93287.1; JOINED.
A PIR; T21371; T21371.
A HSP; P15167; IDTH.
A MEROPS; M12.135; -.
A WormPep; F25H8.3; CE05729.
A GO; GO:004222; F:metalloendopeptidase activity; IEA.
A GO; GO:000270; F:zinc ion binding; IEA.
A GO; GO:006508; P:proteolysis and peptidolysis; IEA.
A InterPro; IPR001590; Peptidase_M12B.
A InterPro; IPR006025; Pept_M_Zn_BS.
A InterPro; IPR000884; TSP1.
A Pfam; PF01421; Reprolysin; 1.
A Pfam; PF00090; tsp 1; 14.
A SMART; SM00209; TSP1; 18.
A PROSITE; PS0215; ADAM_MERO; 1.
A PROSITE; PS00392; TSP1; 15.
A PROSITE; PS00142; ZINC_PROTEASE; 1.
A SEQUENCE 2165 AA; 244397 MW; FCC3DA8AAA9C4888 CRC64;

```

```

Query Match      8.4%; Score 694; DB 5; Length 2165;
Best Local Similarity 22.9%; Pred. No. 4.5e-56;
Matches 351; Conservative 133; Mismatches 556; Indels 492; Gaps 84;

QY 18 CEHPALQPTPEQAACLPAPR-----DCVSEFTLWNCCKGCKKQHRTRAVIAP-PLF 71
Db 586 CHH-----GACVRLAPESLTKIDQWGDWRSWGECSTCGGVQKGLRDCDCKPRN 637
QY 72 GGLQCPNLTES-RACDAPISCPLEGEBEYTFSLKVGPPWCKRLPHELKEINPSGRVTLDFNS 130
Db 638 GGYKVCQQRERYRSCNTQ-ECPNWTQPY-----REVQCS-----EFN- 673
QY 131 DSNERTVFKHQSKAHHSKSWAIEIGYQTRQVSCSTRSDQGNAMLSI-----CLQD 181
Db 674 --NKDIGIQGVASTNTTHWPKYANVAPNERCKLYC-RLSGSAAFYLLRDKVVDGTPCDRN 730
QY 182 SFPLTVQSCIMPKDC-----TSQWSNSPCSKTCSRGSLLPG----- 219
Db 731 GDDICVAGACYPAGCDHQLHSLTRDKCGVCGDDSDCKVVKGTFFNEQGTGGYNVWKIP 790
QY 220 FRGRSRNVKMAIGGKCEPPELLEKACIVGELLOCPRYSWRTSEWKECQVSLLEQ- 278
Db 791 AGSANIDIROGYNMKEDDNYLSLRA--ANGEFLN-----GHFQVSLARQOI 837
QY 279 --QD-----PHMHTGCP-----YCGGGIQTREV---YCAQSVPAALRA 313
Db 838 AFQDTVLEYSGSDAITERINGTPIRSDIYHVLSVGSHPDIDISYMTAEPNA----- 892
QY 314 KEYSRVPVEKAL-----CVGPAP-----LPSQ- 334
Db 893 --VIRPISALYLWRVTDTWTECDRACRGQSQKLMCLDMSTHRQSHDRNCONVLKPKQA 950
QY 335 --LCNIPCTDDCIVSSWSAWGLCIHENCHEPQKGFTRQRHVLMESTCPAGCHPLVE 392
Db 951 TRMCNIDCSTRWITEDVS-----SCSAXCSGQKGRQVSCVRWEGDRQTPASEHLCD 1002
QY 393 --SVPCEDEPMCY-----RWLASSEGI-CFPDHGKGLGHRILKAVCONDRGDEVSSLC- 442
Db 1003 RNSKPSDIASCIYDCSGRKNYGEWTSCTSGSNGKMR--KSYCVDDSNRRVDESLOG 1060
QY 443 -PVPPPPERKSCIPCRMDCVLSEWT--EWSQSCSCKNSDKGKOTRSTILALAGE-- 497
Db 1061 REQKEATERECNIPC-----PRWYGHMSECSRSCDG---GVWRHAQCLDAADRET 1110
QY 498 GKKPCPSQALQEHRLCNDHSCMLQHWETSPWPCSEDTLVLTALNATIGWGEATCGVI 557
Db 1111 HTRECGGAQT-QEH--CNEHAC--TWQFGVWSDCS-----AKCGDGV 1148
QY 558 QTRRVFCVSHVGVQVMTKRCPSDSTRPETVRPCPLCKKDCIVTAFSEWTPCPRMQ- 613
Db 1149 QVRDANCTDHRSLVPEHRLCKMEKIT-KPCH--RESCPKYKLGESQCSVSCDGS 1204
QY 614 -----AGNAT-----VKQSYRIRIIIOEAANGQEC- 639
Db 1205 SRRVSCVSGNGTEVMSLQGTASDRPASHQTCMLG--TCPWRNTDWSACSVSGIGHRER 1263
QY 640 --DTLYEERECEDEVSLC-----PVYRWKPKMSPCILVPESVWQGTGSS 682
Db 1264 TTECIYREQSV-DASFGDGTMPETSQTCHLLPCTSKPSSHWSPC-----S 1308
QY 683 EACGKGLQTEAVSC-----ISDNRSAEMME-CLKQTNGMPLLVQECTVPCRE 729
Db 1309 VTCGSGIQTNRSVCTRGSEGTIVDEYFCDRNTRPRLLKTCCKDTCDPGRVLQK----LQA 1364
QY 730 DCTFTAWS--KFTPCSTNCEATKSRRLQTKSRKKEKQD-----SDLYPLVET 777
Db 1365 DVPPIRWATGFWTACSATCG-----NGTORLLKCEDHVRDLDPDEVCNHLDKEVST 1415
QY 778 ELCPCEDEFISQVGNWSDCLLPGRREPURGLRVQADSKECEGGLRFRVACSDKNGRPV 837
Db 1416 RNCRLRDCSYWMAEWEBCPATCG--THVQQRNVTCVSAEDGGRTILKDVDC-DVQKRPT 1473

```

```
QY 838 DPFSSSGYIQKCVIPCP-FDCKLSDW--SSWGSSSSSCGIVRIRSKWLKEKPYNGG 894
D 1474 SARNCR-----LEPCXGEEHIGSWIGDKSCASCGGWRSSVSCSTSSSCDET 1524
QY 895 RPKCLDLQNVHEAVPCYSCNO-----YSWVVEHWSCKINNELSLRSLCGGQTOS 946
D 1525 RK-PK-----FDKNEELCPPLTNSWQISPHWTHC-----SVSCGGGVQR 1564
QY 947 RKICVNTADGEGVANDNLNODBI-----PPTOSC-----SL 981
D 1565 RKIWCEVDLSGR-----KODDIECSSEIKPREQDCMPPCRSHYHNTKSSASMTSL 1615
QY 982 MCPNECVMS-----EW--GLASKCPQSCDPHTMORTRHLLRPSLN-----S 1021
D 1616 SSSNSNTSSASASLPLPVVSWQTSAMSACAKGRGT-KRRVVECVNPSLNTVTAS 1674
QY 1022 RTCABDSQVPC-----LLNENCFOFQVNLTEWSTQLSENAPQGGVTRTLSCVSDGK 1077
D 1675 TEC--DQTKPVEEVRCRTKEC--PRWKTITWSSCSVT-----CGRGIRREVQCYRGKN 1726
QY 1078 PVMDOCEQHNLEKQPMISIPCLVECV-VNQLSGW--TATWESQTCGHGGRMSRTRI 1134
D 1727 LVSDSECN-----PKTKLNSVANCFFVACPAYRWNVTWMSCKDECAKQKQTRVHC 1779
QY 1135 IMPTQGEGRPCFTLTOKT-----CPVTPC--YSWVLGNWSACKLEGDCGEGVQIRSLSC 1189
D 1780 ISTSGRAAPRMCELARAPTSIRECDTSNCFYEWVPGDWQTC--SKSCGEGVQTRVRC 1836
QY 1190 --MVHSGSISHAAGVEDALCGEMFQDSILKQLCSV-PCPGD---CHLT-----EWS- 1236
D 1837 RKKNFNTSTIPFIWLED-----BPAVPEKCELFPPKNESQTCELNPDCESEFKWSF 1888
QY 1237 -EWSTCELTCT-----IDGRSFETV 1254
D 1889 GPWGECSKNCQGIIRRRVKVANDGRVERV 1920

RESULT 10
QXSV8 PRELIMINARY; PRT; 1637 AA.
AC QXSV8;
DT 01-NOV-1999 (TREMELrel. 12, Created)
DT 01-NOV-1999 (TREMELrel. 12, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE SCO-spondin (fragment).
GN BOS-SPONDIN.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_taxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Gobron S., Creveaux I., Didier R., Meintzel R.;
RT "Characterization of cattle SCO-spondin.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ123488; CAB46239.1; -.
DR HSP; P56882; ICCV.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0005179; F:hormone activity; IEA.
DR InterPro; IPR006208; Cys_knot.
DR InterPro; IPR006207; Cys_knot_C.
DR InterPro; IPR001545; Gly_hormoneb.
DR InterPro; IPR002319; TIL_Cysrich.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR008085; TSP 1.
DR InterPro; IPR001007; VNF_C.
DR Pfam; PF000007; Cys_knot; 1.
DR Pfam; PF01826; TIL; 3.
DR Pfam; PF00090; tsp_1; 11.
DR Pfam; PF00093; vwc_1.
DR PRINTS; PR01705; TSP1REPEAT.
DR SMART; SM00041; CT; 1.
```

```
DR SMART; SM00068; GHB; 1.
DR SMART; SM00209; TSP1; 11.
DR SMART; SM00214; VWC; 2.
DR PROSITE; PS01225; CTCK 2; 1.
DR PROSITE; PS00261; GLYC_HORMONE_BETA_1; 2.
DR PROSITE; PS50092; TSP1_11.
DR PROSITE; PS01208; VWC_1; 1.
DR PROSITE; PS0184; VWC_2; 1.
FT NON_TER 1
SQ SEQUENCE 1637 AA; 173347 MW; 4CSBABIDD346C925 CRC64;

Query Match 8.0%; Score 661; DB 6; Length 1637;
Best Local Similarity 22.1%; Pred. No. 4.3e-53;
Matches 353; Conservative 126; Mismatches 516; Indels 604; Gaps 90;

QY 10 RTVANEICEHFALOPPTTEQACLIPCRDC-VVSEFLPWSNCSKGC-GKKLOHTRAVIA 67
D 269 RTGCNCSAHLKSCSYE-----ACSKAAGFPWGPWPCSRSCGGLGTRTRSRQCVR 323
QY 68 P-PLFGGLQC--PNLTESRACDAPISCPLEGEETFSLKVG-----PWSKRLPHL 115
D 324 PNPAPGGQGHGP-WDLEYCPSP-ECP-GAAGSTAEPATGLPGWGLWSWSPSCGTCT 380
QY 116 KEINPSGRTVLDNSDNERTVTFKHQSKAHHSKSWAIEIGYQTRQVSCT---RSDQN 172
D 381 DPAHPAWR-----SRRLCL-----ANCTGGAASQERP 408
QY 173 AMLSCLQDSPLTVQSCIMPDKCETSONSWSPSCSKTCRSGSLLPGRSSRNVMGMAI 232
D 409 CNLPCTE--LPLCPGCEAGNCSWTAWAPWPCSRSCGVG-----QORRLAYHPG 460
QY 233 GGGKECPPELL-----EKEACIVEGELLQCP-RYSW-RTSEWKBCQVSLLEQDPPHVT 286
D 461 PGHWCPDVLTAQERRFC-----NLRACPVGGSRWSPMSWCDRS-----502
QY 287 GPVCGGIGITREYVCAQSVPAALAAKEVSEPVKE-----ALCVGPAPLPSOLCH-IPCST 342
D 503 -----CGGG-----RSLSRSCSPPPKNGGAPCVGERH-HAKLNTPTCEE 543
QY 343 DCIVSSWSAWGLCIHENCFPGKKGFTQRHVLMESTGPAGHCPH-----LVESVPC-E 397
D 544 GC-----PAG-----NEVVSANCRPCRCSDLQEGIVCOE 573
QY 398 DPMCYRW-----LASGICFP-DHGKC--GLGHRIL-----KAVCON--DRGEDVSGS 440
D 574 DQACQCGRCPEGSLEQDGGCVPLGHCECTDAQGHSWAPSGHQEACNCTCRAGLSCT 633
QY 441 LCPVPPPPPERKSCIEPCRMDCVLSWETWSSCSOSCNKNSDGKQTRSRITILALAGEGK 500
D 634 AQPCPPAH-----CAWSENASWSPCSRSC--GPAGQQRFRS--STSGSWAP 677
QY 501 PCPPSQALQEHRLCNHSCMLHWETSWMGSCSEDTLVTALNATIGNGEATCGVGIQTR 560
D 678 ECRERQSOS-----QPCQSPC-----694
QY 561 RVFCVKSHVQVMTKRCPDSTRTPTVRCFL-----PCKKDCIV-----TA 601
D 695 -----PPLCLQGTFRSLGSLWLDGQCQCSCTPFGIICEDACAGLWATP 741
QY 602 FSEWTPCPRMCOAGNATVQSKRYRIIOEA-ANGQEC--PDTLYEERECEDEVSLCPVY- 657
D 742 WSPSDDCPVSCGGN-----QVTRVCVASAPRGGSPCLGPDV--QSQR-GLMPCPALP 794
QY 658 ---RWKPKWSPCILVPESVMQITGSSEACGKGLQTRAVSCISDNNRSEAMECLK--- 711
D 795 DTCSWGP--WGPC-----SRSCPGGLASRSAC-----PCLLAE 827
QY 712 ----QTNGMPLLVQECTV-PCREDCTFTAWSKFTPCSTNCEATKSRREQ----- 755
D 828 EPACNSTSPDLDTQACYAGPCLEECVWSSWSWTRCS--CEVLVQQRVHQRPAPGGAGA 885
QY 756 -----LTGSKRKEKQDSLDLYPLVETELC--PCDEFISQPYGNW----- 793
```





2798 GVRAR--FRSPSPNPAAMGAPCEGRQELQGGCHTVCGTGIAGISLGAAGVPPSSQFCTLR 2855  
340 C-----SDTCIV-----SSWANGLICIHENCHEPQKKGKGETRQRFVLMESTGPAGHC 387  
2856 THMGPTDHTGTGIEVPGWTFPTWTSWSC--SQSLAPGGGPGWRSR--LCPSPGDSS--C 2911  
388 P-HLVESVPCEDPM-CYRWLASBEGICFPD-----HGK-----417  
2912 PGDQTEPCSPPIECTGFCAPGCTCPGFLFLNASCLPRSCQPCQLHQLYASGAMARL 2971  
418 -----CGLHRLKXAVCONDRGEDVSGSLCVPPPPPKSKCEIFCRMDCVLSEWTS 470  
2972 DSCNNCTCVSG-----KNACTSER-----CPVA-----CGSPWTLWS 3004  
471 SCSSCSNKNKSDGKQTRSRITLA-LAGEGKPCPPSOALQEHRLCHDHC-----M 520  
3005 LCSCSC-----NVGLRRFRAGTAGTAPPAFGAEC--QGPTAEAFCSLRPCGVPVGMCFR 3058  
521 QLHWETSPWGP--CSEDTLVLTALNATI-----545  
3059 DKQLDCAQGPASCALSAFRGNTQCTHPGCHCPGSMGLLLVSPRHPGLGASVQPPVAL 3118  
546 -----GW-----NGEATCGVGIOTRRVFC-----VKSHV 569  
3119 PGATGTSVPGAGGWPWNSHSCSGGGLASRTACDQPPQGLGDYCEGPRAQGEV 3178  
570 QGVNTRKCP--DSTRPETVR--PCFLPKKDCIVTAFSEWTPCPRMCOAGNATVKSRYR 625  
3179 QCALP--CPVTNCTAIBGABYSPCGPPCRSCDDLHVCVWR--QPGCYCPGPGVLSNGA 3235  
626 IICGAAANGGQECPTLYEER-----ECEDVSLCVFYRWKQP 662  
3236 ICVPQ--GHCSCLDLTGTQHHHPGARLAPDGNCHCTCLEGRNCTDLP-CPDC-GGGQ 3290  
663 KWSFC-----ILVPESYVQ--GITGSSEACG-----KGLQTRAVSICSDNRSR 704  
3291 SLHFCGQPCPRSCQDLSPGSCVCPGSCVCOPTCGPLGQLSQDLGVPAHC-----3342  
705 EMMECLQKTMG-PLLYQECV-----PCRE--DC-----TFTANSKFTPCSTNC--EATK 750  
3343 ---RCQYQPGAMAFSPVSTCVAGILOQCEVPDPCDPGVMSWGWFWEDCSVCGGGEQLR 3399  
751 SRR-----RLTKSRKKEKQDSDIYPLVET--ELCPCD-EPISQPYGNMSD 795  
3400 SRRCARPPCPGAPASQRTCSQVCREAGCPAGRLYRECQPGEGCFSCAHTVQVGCFSF 3459  
796 ---CILPEGRREPHRLRVQ-----ADSKCEGEGLR 823  
3460 GCEGCHCEGTFO-HRLACYQECFCVLTAWLLOELGATIGDPQPLGPGDELDSGQTLR 3518  
824 FRVACSDKNGR--PVDPSFCSSGY-----IQKCVIPCFP--858  
3519 TSCGNCSAHGKLSCLDDCEADGGGFWNSPWPSCRSOGLGTRTRESQCVLTWPTLS 3578  
859 -----DCKLSDWSWSGSSSCGIGVIRIKMKLKEKPYNGRCPKLDLUNQ 905  
3579 BLPVCPGPGCAGNCSWTSAFWPFCPSRSCVGGQRLRAYR--PFGPGHWCNMLTAYQ 3637  
906 VH-----EAVPCVSECNQYSWVVEHWSCKINNEL--RSLRCGGGTQSR 947  
3638 ERRFCNLACPEAGCPAGNEVVTANCPRCSDLQEGIVQDDVQVQKGCRCFKGSLQ 3697  
948 KIRCV-----NTADGEG-----GAVDSNLCNDEIPPETOSCSLM--C--PNECVMSWGL 994  
3698 DGGCVPIGHCDCTAQGHSWAPGSGHQACNCSQAGQLSCTAQPCPPPTHCASHWSA 3757  
995 WSKCPQSDPHTMORTTHLLRPSLNSR--TCAEDSQVQPCLL-----NE 1037  
3758 WSPCHSGCPGQGRFR--CGPLASRSGC-----PCLMKADPTCNSTFLHLDTO 3808  
1038 NCF-----QFYQNL--TEWSTCOLSENAPCGQG--VTRLLSCVCSDB--KPVSMDCQEQ 1086  
3809 GCYSGPCPDCQWSLWGPWSPQV-----PCSGGFLRWREAEALCGGSCREPNADR---3861

1087 HNLEKFORMSIP--CLVECVV-----NCOLSG-WTAMTECSOTCGHGRMGRTRFIIMPTQ 1139  
3862 ---XJQRRALPSTCVNESLVCPHQECFVLGPHASWSSCSAPCG--GGTMEHRHC-----EG 3913  
1140 GEG-RPCELT-OKTCTPVTPCYSWVLGN-WSACKLEGGDCGEGVQIRLSQM---VHS 1193  
3914 GPGVAPCAQTEQRCNQLQPCPECPFGQVLSAC-----ATSCPLCWLQLP 3961  
1194 GSISHAAGRVEDALCGMPFODSILKOLCSVPQGDCHLTWSEWSTCELTCDIGRSFET 1253  
3962 G-----ALCVBPCPG-----CG--CPEGGHSILPWGLTILE-----3992  
1254 VGRSRSRTFIQSFENQDSCQVLETRPCTGCKYHYTWKASLMNNERTVWCQSDG 1313  
3993 -----EQAQELPPGTVLNRCTRCVCHGAFSCSL-----VDCQGE--4028  
1314 VNVTCGCSPOARPAIROCIPAC-----RKFYSYCTGGVCGC 1351  
4029 IVPGETWQVAPGELGLCEQTCLMNAATKQSCNSARASGC 4071  
  
RESULT 12  
Q8MYA8 PRELIMINARY; PRT; 1461 AA.  
AC Q8MYA8, Q17591, Q22300;  
DT 01-OCT-2002 (TRENBLrel. 22, Created)  
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE C. elegans ADT-1 protein (Corresponding sequence C02B4.1).  
GN ADT-1.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RN SEQUENCE FROM N.A.  
RA Kuno K., Baba C., Asaka A., Hosono R.;  
RT "Analyses of C. elegans ADAMTS family gene, adt-1";  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RN SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA none;  
RT "Genome sequence of the nematode C. elegans: A platform for  
investigating biology";  
RL Science 282:2012-2018 (1998).  
RN [3]  
RN SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA McMurray A.A.;  
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB066246; BAC05514.1;  
DR EMBL; Z50004; CAA90293.2;  
DR EMBL; Z50006; CAA90293.2; JOINED.  
DR EMBL; Z50004; CAA90302.2; JOINED.  
DR EMBL; Z50006; CAA90302.2;  
DR PIR; T18856; T18856.  
DR GO; GO:0004222; P:metalloendopeptidase activity; IEA.  
DR GO; GO:0008270; P:zinc ion binding; IEA.  
DR GO; GO:0008508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR001590; Peptidase\_M12B.  
DR InterPro; IPR006025; Pept\_M\_Zn\_BS.  
DR InterPro; IPR000884; TSP1.  
DR InterPro; IPR008085; TSP 1.  
DR Pfam; PF01421; Repolysin; 1.  
DR Pfam; PF00090; TSP 1; 12.  
DR PRINTS; PR01705; TSP1REPEAT.  
DR SMART; SM00209; TSP1; 13.  
DR PROSITE; PS00215; ADAM\_MEPPO; 1.  
DR PROSITE; PS00092; TSP1; 12.  
DR PROSITE; PS00142; ZINC\_PROTEASE; 1.

Q SEQUENCE 1461 AA; 162616 MW; 3CFDC1C07C1F493A CRC64;  
Query Match 7.1%; Score 587.5; DB 5; Length 1461;  
Best Local Similarity 23.4%; Pred. No. 3.7e-46;  
Matches 315; Conservative 155; Mismatches 460; Indels 419; Gaps 86;  
Y 12 VVANET-----CEHFALQPTQACILPCPRCIVSEF-----LPWSNCS----- 51  
b 385 VLAHEGHNMGVHDQVQCNKGC-----CLMSAVGAGKTTWSDCSVREFNAFLQ 437  
Y 52 ---KGGCKLQHRTRAVIAPFLPGLQCP--NLTSRACDAPISCPLEBEEYTFELKVP 106  
b 438 LBESGEGNCLRASGLISTNHLSDLRLPQGFADQOC-----SYFWR 482  
Y 107 WSKRLPHLKEINPSGRTVLDNFNSDNEXVTFKQSYKAHHHSKWAIEIGVQTRQVST 166  
b 483 DYKVEIPNGKAMDDICRIWCGNSGST--ISTAHPALEG-----SWC----- 522  
Y 167 RSDGQNAL--SLCLQDSPLTVQSCIMPKDCETQWSWSPCSKTCR-----SGSLLP 218  
b 523 ---GANKWCHKQCTHTWTFGLTP-----VFID-----GEWSEWGAEGKCPICQCAVSGSITV 572  
Y 219 GFRSRNRNKHMAIGGKCEPALLEKEAC-----IVEGELLOQCPRYSWRTSEWKECQVSL 274  
b 573 QGHRD--CVNPAPNNGKTC-----EGANIRGIVCGATSSNCLGTFREEFGNKICS--SI 624  
Y 275 LLEQDPHHVTVGPVCGGIGIOTREYVCAQSVFAAALRAKEVSRPVEKALCVGPAPLPQ 334  
b 625 KYDPHPDQOLTEGEFEHTQPCRVWC--HLTSELIRNG----- 663  
Y 335 LCNIPGTCIVSSWSAMGLCIHENHEPQGGKPRTRQHRVLMESTGPAGCHPLVESV 394  
b 664 ---QFPDGTGCGDAVCGQCLALSC-----DNKALVEQ-----PEDCPRI----- 702  
Y 395 PCEDPWCYRW--LABEGICFPDHGKGLGHR--ILKAVQNDRG--EDVSGSLCPVPPPPER 450  
b 703 ---EGRSVHWEWSSWSEC-----SVSGGLGREGVREKCSGKQCVS-----EES 748  
Y 451 KSCETPCRMDC--VLSEWTWSSCSCSKNKGDKQTSRTILALAGGKPCPPSQ-- 507  
b 749 RCEGVLR--DCEFEWKEWGSCEKAL-----GVQKPR-----PCLTDQCS 792  
Y 508 ---LOEHLNDRHSCMLHWETSPWGPCSDETLVTALNATIGNMGATCGVGIOTRRVFCV 565  
b 793 KHLQERPCDNEGCV--TNWD--EWSSCSQ-----SCGGRRYR----- 827  
Y 566 KSHVQVMTKCP--DSTRPETVRPCLPCKDCIVTAFSEWTPCPRMQAGNATVQSBRY 624  
b 828 ---IRKCLDDKCDGDDLEKESCN-----TQKCSQSGMDMLPCSVSCGIG----- 869  
Y 625 RIIOEAANGGQECPTLYEERECEDEVSLCPVYRWKPKWSPCLVPSVWQGITGSSEA 684  
b 870 -FQIREELCDGELCATANKQARTCQ--QOCP-----SAFSLVSWSEWETTCAT 918  
Y 685 CKGLOT-----RAVSCIDDNRSAEMELCKQTGMELLVQECTVPCREDCTFTAWSKF 739  
b 919 CGEGLRSRSCRRGCTEDD--ASQTRCV--NG-----PCEH--SVLTWSEW 961  
Y 740 TPCSTNCEATKRRRLTCKSRKKEKQCDSDLYPVVEVELCPCEDEFISQPYGNWSDCILP 799  
b 962 TTCET--CSSFDRKR--IAKCDGTENCQK-----IDETCDI-----ACL-- 1000  
Y 800 EGRPHRGLRVOADKEGEGELRFAVACSKN---GRPVDPSFCSSGVIQEKVIPC 856  
b 1001 ---REKHSFGPISP-----RRPLITSNLRKAFGRPLLPISHS-----EK----- 1040  
Y 857 PFDCKLSDWSSGSCSSCGIGVIRSKWLEKPKYNGGRPCPKLDKNQVHEAVPCYSEC 916  
b 1041 -----WSEWGPCSVTCGSRVRTRCQE-----ASCPEHQIQT-----EC 1077  
Y 917 NOYS-----WVFEWSSCKINNELSLRCG--GGTQSRKIRCV--NTADGGGAVDSNLCNQD 970  
b 1078 NLNSCLELFIWSDWSSC-----SKSCQGDGIQTRQK--CLFNNAECSSYA--ESRCK-- 1127

QY 971 EIPPETQSCSLMCPNECV-----MSEWGLWSKPOSCDPHT--MQRTRHLLRPSLN 1020  
Db 1128 DLP-----SCSASSIGRTISENGFDPARWSEWSKSAC--SCFSLTSTRRFFCQVDPVTVQ 1181  
QY 1021 SRCAEDSVQVQPCLLNENCFOFO--YNL--TENWTCQLESNAPCGQVTRLLSCVSDGK 1077  
Db 1182 GFCAGALEIQCPCAPG--SCSPSAGHSLSEWSSCS--KDCGDTGHQIRNM-----CSEPI 1235  
QY 1078 PVSMD--QCEOHNELEKFORMSIPCLVECV-----VNCQLSGWTAWTECSOTCGHGRMSRT 1131  
Db 1236 PSNRGAYCSYSD--QR---PCVMNVCSDEKVDGGWTDWTAMSECTDYCRNGHR--SRT 1289  
QY 1132 RFIIMPQGEGRCPTELTOEKTCPVTPCY-----SWVLGNWSACKLGGDC 1178  
Db 1290 RFCANPKPSQGAQCTGSDPE-----LNPFCDPARCHLRDGGWSTW--SDWTPC--SASC 1340  
QY 1179 GEGVQIRLSLSCMVHSGSISHAAGRVEDALCGEMPFODSILKQLCSVPC---PGDCHLW 1235  
Db 1341 GFGVQTRDRSC-----SSPEPKGQ-----SCSGLAHQTS-----LCDLPACDHESDGEWSAW 1388  
QY 1236 SEWSTCELTICIDGRSPETVGRQSRRTFI 1264  
Db 1389 NEWSGCMGNC-----GIGTRTRVRACV 1410  
RESULT 13  
O43376 PRELIMINARY; PRT; 182 AA.  
AC O43376;  
DT 01-JUN-1998 (TEMBLrel. 06, Created)  
DT 01-JUN-1998 (TEMBLrel. 06, Last sequence update)  
DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)  
DE Hypothetical protein D1110N13.1 in chromosome 7 (Fragment).  
GN D1110N13.1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Andrews S., Holmes A.;  
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC004141; AAC02753.1;  
DR InterPro; IPR000884; TSP1.  
DR Pfam; PF00030; TSP\_1; 1.  
DR SMART; SM00209; TSP1; 1.  
DR PROSITE; PSS0092; TSP1; 1.  
KW Hypothetical protein.  
FT NON TER 1  
FT NON TER 182 182  
SQ SEQUENCE 182 AA; 20189 MW; B898DF494381AB71 CRC64;  
Query Match 6.7%; Score 555.5; DB 4; Length 182;  
Best Local Similarity 52.5%; Pred. NO. 2.5e-44;  
Matches 95; Conservative 29; Mismatches 56; Indels 1; Gaps 1;  
QY 577 CPDSTRPETVRPCLPCKDCIVTAFSEWTPCPRMQAGNATV--KQSYRIIITQEAANG 635  
Db 1 CPESLRETVRPCLLCKKDCIVTSDWTSCPSCKEGSSIRKQSRHRIQLPANG 60  
QY 636 QCEPDTLYERECEDEVSLCPVYRWKPKWSPCLVPSVWQGITGSSEACGKLOTRAVS 695  
Db 61 RDCDTPLYEKACEAPQACQSYRWKTHKWRRCQLVPSVWQGITGSSEACGKLOTRAVS 120  
QY 696 CISEDNRASMMELCKQTGMELLVQECTVPCREDCTFTAWSKFTPCSTNCEATKRRRQ 755  
Db 121 CRQDGGQAGIHCLQVAGVPVPAQTACQIPCDQCQLTSWSEKSSCGDCGAVTRKRT 180  
QY 756 L 756  
Db 181 L 181

## RESULT 14

Q8T3A0 PRELIMINARY; PRT: 1089 AA.  
 ID Q8T3A0  
 AC Q8T3A0; 21, Created)  
 DT 01-JUN-2002 (TREMUREL. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMUREL. 21, Last sequence update)  
 DT 01-OCT-2003 (TREMUREL. 25, Last annotation update)  
 DE Putative coagulation serine protease.  
 GN SP4  
 OS Ciona intestinalis.  
 OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;  
 OC Phlebobranchia; Cionidae; Ciona.  
 OX NCBI\_TaxID=7719;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC TISSUE=Hepatopancreas;  
 RA Hammond J.A., Nakao M., Yano T., Kemp G.D., Smith V.J.;  
 RT "Complete cDNA sequence of a serine protease from Ciona  
 intestinalis.";  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ431686; CAD24309.1; -;  
 DR HSSP; P00761; IAN1.  
 DR GO; GO:0016020; C-membrane; IEA.  
 DR GO; GO:0005509; F:calcium ion binding; IEA.  
 DR GO; GO:0004263; F:chymotrypsin activity; IEA.  
 DR GO; GO:0008233; F:peptidase activity; IEA.  
 DR GO; GO:0005044; F:scavenger receptor activity; IEA.  
 DR GO; GO:0004295; F:trypsin activity; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR009003; Cys\_Ser\_trypsin.  
 DR InterPro; IPR001881; EGF\_Ca.  
 DR InterPro; IPR006209; EGF-like.  
 DR InterPro; IPR006210; IEGF.  
 DR InterPro; IPR001254; Peptidase\_S1.  
 DR InterPro; IPR001314; Peptidase\_S1A.  
 DR InterPro; IPR003582; SHKT.  
 DR InterPro; IPR001190; Srcr\_receptor.  
 DR InterPro; IPR008884; TSPI.  
 DR InterPro; IPR008085; TSP\_1.  
 DR Pfam; PF01549; SHKT; 3.  
 DR Pfam; PF00089; trypsin; 1.  
 DR Pfam; PF00090; tsp\_1; 11.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR PRINTS; PR01705; TSP1REPEAT.  
 DR SMART; SM00181; EGF; 1.  
 DR SMART; SM00254; SHKT; 3.  
 DR SMART; SM00020; Tryp\_Spc; 1.  
 DR SMART; SM00209; TSPI; 11.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS01187; EGF\_CA; 1.  
 DR PROSITE; PS0287; SRCR\_2; 1.  
 DR PROSITE; PS0240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 DR PROSITE; PS00092; TSPI; 11.  
 DR EGF-like domain; Hydrolyase; Protease; Serine protease.  
 SQ SEQUENCE 1089 AA; 113693 MW; 2742DE15E22F6C0A CRC64;

Query Match 5.8%; Score 474; DB 5; Length 1089;  
 Best Local Similarity 20.7%; Pred. No. 1.6e-35;  
 Matches 275; Conservative 117; Mismatches 425; Indels 510; Gaps 73;  
 Qy 168 SDGQAMLSLCLQDSFFLTVQSCI-----MPKDCETQSWSPSCSKTCSRGSLLPG 219  
 Db 21 SQGQSSVCSTSL-----GCVDCFSWCQANAASCTSSPALMGSYCKXTCN----- 64  
 Qy 220 FRSRNRVKNVAIGGKECPLELKE---ACIVGELLQOCPRYSWTSWKE-----CQV 272  
 Db 65 -LCASNSAACAKGNCNHCRIETTSSEPVACAF-EGFRLEANGRTCTVDIDEAENSTLCS 122  
 Qy 273 SLLEQDPHHVHTGPGVCGGIGIQTRVY-----CAQSVPAAL-----RAKEY 316

Db 123 PNLNMCNNTLGHVYCTACGSPNHRHAYVERNECKMGSGACGKSTNGGRIVGKGRGRIA 182  
 Qy 317 SRP-----VZKALCVGPA-----PLPSQLC-----NIPSTDICV----- 346  
 Db 183 RWPWMAYIVIGRNLCGGTLLSSGWLTAAHCFASITNNNPSTINNVILGVDTDSGNIHE 242  
 Qy 347 SWSAWGLCIHENCHEPOGKGFRTQRHVLMESTGPAGHC- PHLVESVPCEDEPMC----- 401  
 Db 243 QSFVSVTRLIIHPNPNNDLALLQDLHDALIDAAFKVPCVCLPNGEE--PPEGKCKWATG 300  
 Qy 402 YRWLASEGIC-----FPDHGKC-----GLGHRI-----LKAVCONDR 433  
 Db 301 YGTIAFGVAAKSIQEVNDLPADLAHCERIYANTNRVNTMLCAGYITGCKDTCQDS 360  
 Qy 434 GEDVSGSLCPVPPPPERKSC-----IPCRM----- 459  
 Db 361 GGPLVCQRC-----KNCDWYLAGITTSFGRGCARPGFFGVYTKVSFFEQWISSYTSIA 412  
 Qy 460 ----DCVLSEWTEWSS-----CSQSCSNKNSDKQTRSRITLALAGEGKP-CPPSQALQEH 511  
 Db 413 INPGQCVKPSWTTWGSWTPCA-SCSGSS-----SRIRFCANGSPDGPDCGLQ--EEF 462  
 Qy 512 RLCNDHSCWQLHW-ETSPWGPCSEDTLVTALNATIGNGEATCGVGIQTRRVFVKSHV- 569  
 Db 463 ROCST-VCTQPTWAEYGDWGS-----VTCGDGSRSRICENGNI 504  
 Qy 570 -----QGVMTK-----RCPD----- 579  
 Db 505 DRPGCTGGETATEACTTGVRCPTWSAMSGYGVCSVTCGGGTQESTRTCNHNGQAGVTC 564  
 Qy 580 -STRPETVRPCFLPCKKDCIVTAFSEWTPCPRMCOAGNATVQSKRYRIIQEAANGQEC 638  
 Db 565 RDRSQACNPOTCFAP---TWAAVCAWSDCTRCGGGERT-----RVETCL-NGAIGSGC 616  
 Qy 639 PDLTYERECEEDVSLCPVY-RWKPQ-KWSPCILVPESVMQGITSSBACGKLOTRAVSC 696  
 Db 617 PAAGVSQTESCNIOSCQANPTWSAYGWSGC-----SVTCASGTRTSRSC 662  
 Qy 697 ISDD--NRSAEMWECLKQTNMPLLVQECT--VPCREDCTFTAKSKTTPCSTNC-----E 747  
 Db 663 VGNIGNVGCE-----SGQTAASEACTTGVQCP---TWSAMSVYGVCSVTCGGGTQ 711  
 Qy 748 ATKSRRRQ-----LTGSKRKEKQSDSLYVELTELCPCEDEFISQPYGNWSDCILPEG 801  
 Db 712 STRTCNNHGVGVTCGDRDTRSQAC-----NPQACPS-----WSGYGWSGC----- 753  
 Qy 802 RREPHRGLRVQADSKCEGGLRFRVAVACDKNGRPVPSFCSSSGYICEK---CVIPCPF 858  
 Db 754 -----SETCGDGTKTTRTCN--NGQ-IGDNGCSPAAATDMSMACSVRNCP- 796  
 Qy 859 DCKLSDWSWSGSCSSCGIGVIRSKML-----KEKPYNGRCPCKL 902  
 Db 797 --QWSSWSGWSGQCSLTGCGGRTAVRCNTFGATCASCAGATSKSEFCNLG----- 846  
 Qy 903 KQVHEAVPCISECNQYSWVHEVHSSKINNELSLRCGGGTQSKRKRCVNTADGEGAV 962  
 Db 847 -----ACPFVS-----AW--SGWSTC-----SAGCGGQQTTRTCSFGNCDP 886  
 Qy 963 DSNLNCQDEIPEPTOSCLMCPNECMSEWGLW--SKPOSQDPHTMQRTRLLRPSL 1019  
 Db 887 GTAL-----SGSQACNTDAGTGVWNSGTCSAACGPTIQ----- 925  
 Qy 1020 NSRTCAEDSQVQPCLLNENCFOYNLTWSTCOLSENAPCGQGVTRLLSCVSDGKPV 1079  
 Db 926 -TRECIGGTAGP-----NC-----VGSTQQTAAACN----- 950  
 Qy 1080 SMDQCEQHNLEKQPMWSIPCLVECVNQLSGWTAWTECSOTCGHGRMSRTRFLIMPTQ 1139  
 Db 951 -----VAACWGEVWNTACTVTCG-AGTQTRSR-----TCS 981  
 Qy 1140 GREGPCP---TELTOEKTCPVTTPCVSYWLGNSACKLEGGDCGEGVQIR-SLSCMWSGS 1195  
 Db 982 GEAGRCPGQSAATESQACAASTCASTV-----NDCSNDIDLAPAITCREVA-- 1028

```

2Y 1196 ISHAARVE-----DALC-----GEMP-----PQDSILKQLCSVCPQGDCHL--TWSEW 1238
2b 1029 ---VAGCEQYKDYMDINCIRSCVGRNFCSMYRDLFLQ-----CPSYRLHCLNTLVE 1079

2Y 1239 STCELC 1245
2b 1080 PLCKYTC 1086

RESULT 15
2966P9
ID Q966P9 PRELIMINARY; PRT; 1121 AA.
AC Q966P9;
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Hypothetical protein C36B7.5a.
CN C36B7.5.
CS Caenorhabditis elegans.
CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
CC Rhabditidae; Peloderinae; Caenorhabditis.
CX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RL "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.,"
RL Science 282:1202-1218(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Minx P.; Minx M.;
RL "The sequence of C. elegans cosmid C36B7.,"
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL "Direct submission.,"
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006615; AAK68231.1; -
DR WormPep; C36B7.5a; CE25806.
DR InterPro; IPR003014; PAN.
DR Pfam; PF00024; PAN; 1.
DR Pfam; PF00090; tsp.1; 9.
DR SMART; SM00209; TSP1; 11.
DR PROSITE; PS00092; TSP1; 10.
SQ SEQUENCE 1121 AA; 126159 MW; 151B03F47E5BF1BC CRC64;

Query March 5.7%; Score 470; DB 5; Length 1121;
Best Local Similarity 23.0%; Pred. No. 3.9e-35;
Matches 255; Conservative 88; Mismatches 371; Indels 394; Gaps 62;

2Y 357 HENCHE-PQCKGFRTRQHV-----MSTGPAHCPLHVESVPCEDPMCYRWLASGEI 410
2b 197 HQSHEDRGDDTDQRDRYLXITIDYTEKSERPIHPTIHRGQQSSQMSASPVASGNP 256

411 CFPDHGKGLGHLKXAVQNDRGEDVSGSLCPVPPPERKSCIEPC-RMDCV-----462
2b 257 CLFNAACARTLYQMDTAPCPARLGD-----PCAPKE-----PCWDDCYHVAPPQ 302

463 -----LSEWTWSSCSQSCSNKNSDQKTSRTILALAGEGKPCPPSQALQHRCLNDHS 518
2b 303 AAPVVEWTEWSECSASC-----GISLRQCL-----GGITCIGFSTIP-----CQVPE 347

519 CMLHNET-SPWGPCSEDITLVLTALNATIGNWGATCGVG-IQTRRYFCVKSHVGQVMYKR 576
2b 348 CAV--WTLMSFWSLCT-----ATCGVGIEIQRNV-----CQTRG 379

```

Search completed: February 24, 2004, 01:07:25  
 Job time : 142.308 secs

```

QY 577 CPDSTRBETVRC--FLPCKKDCIIVTAFSEWTPCPRMCOAGNATVKOSRYRIIIQEAANG 634
Db 380 NCDGFSVE-VEACKALLPCP---TWSSMWSWTGCGNSC--GKSIERRSR-----ICQN 426

QY 635 QGECPDITLYEERECEDVSLCFVY--RWKPKQWSPCILVPESWQOITGSSBACGKLQTRA 693
Db 427 GLLCFGPATEERTC-DKXPCPHWSNW--GQWKC-----SKECGAGHYRN 469

QY 694 VSCI-----SDDNRSAMMECLKQTNMPLLVQECVTVPREDCTFTAMSKFTPCSTNC--EA 748
Db 470 RECIDGVSCGASEKILCNQO-----PCPEWSQMTAW--TVCDERCGEES 513

QY 749 TKSRRRLQTKSRKKEKQDSLDLYPLVETELC--PCDEFILSQV-----GNWSDCI 797
Db 514 IRLNR-----KQMAE-----NNACDQFADQMSCPYRDCPKWEEWGEWADC- 557

QY 798 LPEGRREPHRLGRVQADSKCEGGLRFRVACSKNGRPVDPSPFCSSSGYIQEKC-VIPC 856
Db 558 -----STTCGQGTQRLRKCDSGNE-----CSGFGEMRFCQIASC 593

QY 857 PFDCKLSDMSSWGSSSSCGIGV--RIR-----R-----882
Db 594 PY---WGDWTPWSCGCVSCGGQGVCEPTRKCIITDFLQPLTEELERDDSLKHEAKBALI 650

QY 883 -----SKW-----LKEKPYNGRCPCKLDLKNQVHEAVPCYSECNQYSWVV 923
Db 651 ARAKTISKYRRTNETRLSPRPPLSPQLGG--TCDGPEMETKVCADGCP---CYVNAWT- 705

QY 924 EHWSC-KINNELSLR---CGGQTQS-----RKIRCVNTAD-GEAGVDSNLCN 968
Db 706 -EWSFCICGCGNEGISKRNVCSEMGQSPAPPPIISFQVRHRTYVTASLGQGPIGNILPG 764

QY 969 QDEIIPP-----ETQSCSLMCPNE-----CVMSSWGL 994
Db 765 MSPIVPVEIHRGKQVLFQIATPKCH--CPGDTFQTRPCLEQNAQODPVGKCEWSWGE 822

QY 995 WSKCPQSCDPTHMQRTRHLLR-----PSLNSRTC-----AEDSQVQPCLLNENCQFQ 1043
Db 823 WCGC-MRCRPGKVR--RFDPSPIGRGGPMRPDSTCDGNGDDSDQERSCPMEQTCYNG 880

QY 1044 YNLTEWSTCQLSENAPCGQGVRTLLSCVCS-----DGKPVSMDCQCHNLEKQRMNSIP 1098
Db 881 IGTGRFNNAASMQDARERQNRGTGTFYNTAAPQDRRTSYNSNRRPSSHRSQDEYVID 940

QY 1099 CLVECVVNCQLSGHTAWTECSQTCGHGRMSRTRFII-----MPTQGEGRPC 1145
Db 941 SVIGRVQVCHWSKWSWSRCHD-----NSTREKRFVYGEKSELVSNCECLGKPEHEEPC 996

QY 1146 PT----ELTQEKTCPTVPCYSWVLGNWSACKLEGDCGEGVQIRSLSCMTHSGSISHAAG 1201
Db 997 NATGFIEQSNBEET-----DREIENNLDKLLBEDNAGGDVELHKKKEELKS-KVESAIG 1048

QY 1202 RVEDALGCEMPFQDSILKQLCSVPCGDCHLTWSEWSTCELTICIDGRSPETVGRQSRSR 1261
Db 1049 TVQ-----CDWTRWSQWSVCTASCGEGR-----RWERRR 1077

QY 1262 TFIQSPENQSCPQOVLETRPCTGGKC 1289
Db 1078 -----C-----PCGESKC 1085

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

un on: February 24, 2004, 01:00:01 ; Search time 53.7866 seconds  
(without alignments)  
1406.151 Million cell updates/sec

itle: US-10-022-710-2

erfect score: 8241

equence: 1 MYRCIQKLNRTVANEICEH.....QSTPPQKPLTAYDGLDM 1465

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

earched: 389414 seqs, 51625971 residues

otal number of hits satisfying chosen parameters: 389414

imum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase : Issued Patents AA.\*

- 1: /cgm2\_6/ptodata/2/iaa/5A\_COMB.pep.\*
- 2: /cgm2\_6/ptodata/2/iaa/5B\_COMB.pep.\*
- 3: /cgm2\_6/ptodata/2/iaa/6A\_COMB.pep.\*
- 4: /cgm2\_6/ptodata/2/iaa/6B\_COMB.pep.\*
- 5: /cgm2\_6/ptodata/2/iaa/PTCUS\_COMB.pep.\*
- 6: /cgm2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

result No.	Score	Query Match	ID	Description
1	694	8.4	US-09-800-729-155	Sequence 155, Appl
2	657	8.0	US-09-369-364A-13	Sequence 13, Appl
3	477.5	5.8	US-09-800-729-89	Sequence 89, Appl
4	440.5	5.3	US-09-132-769-3	Sequence 3, Appli
5	434.5	5.3	US-09-132-769-1	Sequence 1, Appli
6	434.5	5.3	US-09-640-173-186	Sequence 186, App
7	434.5	5.3	US-09-713-550-186	Sequence 186, App
8	429	5.2	US-09-800-729-217	Sequence 217, App
9	428.5	5.2	US-07-862-021B-10	Sequence 10, Appl
10	428.5	5.2	US-08-313-288B-10	Sequence 10, Appl
11	428.5	5.2	US-09-132-769-5	Sequence 5, Appli
12	428.5	5.2	PCT-US93-03164-10	Sequence 10, Appl
13	419	5.1	US-07-862-021B-12	Sequence 12, Appl
14	419	5.1	US-08-313-288B-12	Sequence 12, Appl
15	419	5.1	PCT-US93-03164-12	Sequence 12, Appl
16	383.5	4.7	US-07-862-021B-14	Sequence 14, Appl
17	383.5	4.7	PCT-US93-03164-14	Sequence 14, Appl
18	340.5	4.1	US-08-918-914-4	Sequence 4, Appli
19	330	4.0	US-09-369-364A-17	Sequence 17, Appl
20	329	4.0	US-08-985-526-3	Sequence 3, Appli
21	311.5	3.8	US-08-313-288B-15	Sequence 15, Appl
22	310.5	3.8	US-09-230-682-2	Sequence 2, Appli
23	306	3.7	US-09-930-872-4	Sequence 4, Appli
24	300	3.6	US-08-185-432-19	Sequence 19, Appl
25	300	3.6	US-08-899-232-4	Sequence 4, Appli
26	286	3.5	US-08-083-590A-20	Sequence 20, Appl
27	286	3.5	US-08-532-384-20	Sequence 20, Appl

28	284	3.4	450	4	US-09-369-364A-19	Sequence 19, Appl
29	284	3.4	2556	1	US-08-185-432-17	Sequence 17, Appl
30	284	3.4	2556	1	US-08-899-232-2	Sequence 2, Appli
31	269	3.3	1170	1	US-08-313-288B-20	Sequence 20, Appl
32	264	3.2	2523	1	US-08-185-432-18	Sequence 18, Appl
33	264	3.2	2523	4	US-08-899-232-3	Sequence 3, Appli
34	261	3.2	2471	1	US-08-185-432-16	Sequence 16, Appl
35	261	3.2	2471	1	US-08-083-590A-19	Sequence 19, Appl
36	261	3.2	2471	3	US-08-532-384-19	Sequence 19, Appl
37	261	3.2	2471	4	US-08-899-232-1	Sequence 1, Appli
38	251	3.0	1964	4	US-09-467-997-1	Sequence 1, Appli
39	250	3.0	525	4	US-09-369-364A-21	Sequence 21, Appl
40	240.5	2.9	1205	4	US-09-491-522-11	Sequence 11, Appl
41	239.5	2.9	1251	5	PCT-US95-02251-3	Sequence 3, Appli
42	239.5	2.9	1252	1	US-08-199-780-3	Sequence 3, Appli
43	239.5	2.9	1252	4	US-08-316-650-3	Sequence 3, Appli
44	239	2.9	1211	4	US-09-491-522-5	Sequence 5, Appli
45	232	2.8	1833	3	US-08-479-722B-2	Sequence 2, Appli

## ALIGNMENTS

### RESULT 1

US-09-800-729-155

; Sequence 155, Application US/09800729

; Patent No. 6605592

; GENERAL INFORMATION:

; APPLICANT: Ni et al.

; TITLE OF INVENTION: 32 Human secreted proteins

; FILE REFERENCE: P2044P1

; CURRENT APPLICATION NUMBER: US/09/800,729

; CURRENT FILING DATE: 2001-03-08

; PRIOR APPLICATION NUMBER: PCT/US00/26013

; PRIOR FILING DATE: 2000-09-22

; PRIOR APPLICATION NUMBER: 60/155,709

; PRIOR FILING DATE: 1999-09-24

; NUMBER OF SEQ ID NOS: 217

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 155

; LENGTH: 2165

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-800-729-155

Query Match 8.4%; Score 694; DB 4; Length 2165;

Best Local Similarity 22.9%; Pred.No. 4.1e-47; Indels 492; Gaps 84;  
Matches 351; Conservative 133; Mismatches 556;

Qy	18	CEHALQPPTEQACLI PCPR-----DCVVFSEFLPWSNCSKGGKGLQHRTRAVIAP-PLF 71	
Db	586	CHI-----GACVRLAPESLT KIDGQWDRSNGECSTCGGVQKGLRDCDSPKPRN 637	
Qy	72	GGLOCNLTES-RACDAPISCLGSEETFTSLKVPWPKRLPHLKEINPSTRTVLDNFS 130	
Db	638	GGKYCVGQERYVSCNTQ-ECFWDTPQY-----REVQCS-----EFN- 673	
Qy	131	DSNERTVFHQSKYKAHHSKSWAIEGYOTROVSCTRSDGQNALSL-----CLQD 181	
Db	674	--NKDIGQGVASTNTHWPKYANVAPNERKLYC-RLSGSAFVLLLRKVVGDGTPCDEN 730	
Qy	182	SFPLTVQSCIMPXDCS-----TSQSSWSNPSCKTCRSGSLLPF----- 219	
Db	731	GDDICVAGACMPAGCDHQLHSLRRDKCGVGDDSSCKVKGTFNEQGTFGYNEVMKIP 790	
Qy	220	FRSRNRNVHVAIGGKCEPELLEKACIVEGELLQCCPRYSWRTSEKCECVSLLEQ- 278	
Db	791	AGSANIDIRQKGNKNNKEDDNLSLRA--ANGEFLLN-----GHFQVSLARQOI 837	
Qy	279	--QD-----PHWHTGP-----VCGGGIQTREV---YCAQSVFAAALRA 313	
Db	838	AFQDTVLEYSGSDAIIERNGTPIRSYIVHVLSVGSHPDDISYEYMTAAVENA----- 892	



314 KEVSRPVKAL-----CVGPAP-----LPSO- 334  
893 --VIRISSALYLWVTDTECDRACRGOOSQKLMCLDMSTHRQSHDRNCQWLKPKQA 950  
335 --LCNIPSCDVCISWSAWGLCHENCHPEQKGFRTQRHVLHMESTGPAGHCPHLVE 392  
951 TRMNCIDCSTRWITEDVS-----SCSAKSGGQKQORVSCVMEGDRQTPASEHLCD 1002  
393 --SVPCEDPMCY-----RWLASEGI-CRPDHCKGLGHRILKAVCONDRGSDVSSLC- 442  
1003 RNSKPSDIASCYDCSGRKANYGWTSCSETCSNGKWHR--KSYCVDSNRVRDESLOG 1060  
443 -PVPPPERKSCBPCRMDCVLSEWT--EWSQSCSNKNSDKGKTRSRITLALAGE-- 497  
1061 REQKEATRECNRIFC-----PRWYGVHSECSRSCDG--GVKORHAQCLDAADRET 1110  
498 GGRPCPSQALQEHRLCNHDSCHQLHETSMPWPCSEDTLVTALNATIGNWGEATCGVGI 557  
1111 HTRCGPAQT-QEH--CNEHAC--TWQFGVWSDCS-----AKCGDGV 1148  
558 QTRRVFCVSHVGVMTKRCPSDSTRPETVRPCFLPKCKDCIVTAFSEWTFPCRMCO- 613  
1149 QYRDANCTDRHSVLPBHRCLKMEKIIT-KPCH--RESCPKVKGWCSQSVSCEDGWS 1204  
614 -----AGNAT-----VKQSYRIIIEAANGQCEP----- 639  
1205 SRRVSCVNGNGTEVMSLGTASDRPASHQTCNLG-TCPFWRNTDWSACSVCSCGIGHRR 1263  
640 --DTLYBERCEDVSLC-----PVYRWKQKQSPCLVPSVWQGITSS 682  
1264 TTECIYREQSV-DASFCGDKMPETSQTCHELLPCTSKPWSHNSPC-----S 1308  
683 EACGKGLQTRAVSC-----ISDNRSAEMWE-CLKQTNMGMLLVQECTVPCRE 729  
1309 VTCGSGIQTBSVSTRGSEGTIVDEYFCDRTNRPRLKTCEDKTCDDGPRVLQK-LOA 1364  
730 DCUFTAMS--KFTPCSTNCEATKRRRLQKSKKEKCD-SDLYPLVET 777  
1365 DVPPRWATGFWTACATCG-----NGTORLLKCDHYRDLDPDECNHLDEVST 1415  
778 ELPCDFEISQYGNWSDCILPBGREHFRGLRVQADSKCEGGLRPAVACSKDKNGRPV 937  
1416 RNCRLRDCSYWKAWECEPATCG-THVQOSRNVTCVSAEDGGRTILKDVDC-DVQKRET 1473  
838 DPSPCSSSGVIOEKVCPICP-FDCKLSDM--SSWGSQSSCGIGVIRSKWLKEKPYNG 894  
1474 SARNCR-----LEPCGEBHIGSWIGDWSKCSASCSSGGRWRRSVCTSSCDET 1524  
895 RPKPLDLKNQVHEAVPCYSECNO-----YSWVVEHSSCKINNELASLRCGGGTQS 946  
1525 RK-PKM-----FDKNEELCPPLTNNWQISPWTHC-----SVSCGGGVOR 1564  
947 RKIRCVNTADGEGAVDSNLNDEI-----PPTQSC-----SL 981  
1565 RKIWCEBVLGR-----KQDDISCEIKPREQDCMPPCRSHYHNKTSASMTSL 1615  
982 MCPNECVMS-----EW--GLMSKCPQSCDPTMTORTRHLRLPSN-----S 1021  
1616 SSSNSNTTSSASASSLPIPVVSWQTSAMSACSACKGRGT-KRVVECVNPSLNVTVAS 1674  
1022 RTCAEDSQVQPC-----LLNENCFOFQNLTEWSTCQSENAPCGQVYRTELLSCVSDGX 1077  
1675 TEC--DOTKPFVEVRCTKHC--FRWKTITWSSCSVT-----CGRIRREVQCVRGN 1726  
1078 PVSMDQCEQHNLEKQPMISPLIVECV-VNQCQLSGW--TAWTECSTQTCGGGRMSRTRFI 1134  
1727 LVSPSECN-----PKTKLSNVANCFPVACPAVARNVTWMSCKBECARGQKQTRVHC 1779  
1135 IMPQGGRCPCPTBLTQEK-----CPVTPC-YSWVLGNWSACKLEGDCGEGVQIRSLSC 1189  
1780 ISTSGKAAPRMCELABAPTSIRSCDTSNCPYEWVPGDWQTC--SKSCGEGVQTRVRC 1836  
1190 --MYHSGSISHAAGREVDALCGEMPFQDSILKQICSV-PCFGD--CHLT-----EWS- 1236

1837 RKXNFNSTIPIIMLED-----EPAVPKCKELFKPNESQTCELNPCDSEFKWSF 1888  
1237 -EWSTCELT-----IDGRSPETV 1254  
1889 GPWGECSKNCQGGIRRRRVKVCVANDGRVERV 1920  
RESULT 2  
US-09-369-364A-13  
; Sequence 13, Application US/09369364A  
; Patent No. 6391810  
; GENERAL INFORMATION:  
; APPLICANT: Apte, Suneel  
; APPLICANT: Hurskainen, Tiina L.  
; APPLICANT: Hirohata, Satoshi  
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases  
; FILE REFERENCE: 26473/4007/10-30-00  
; CURRENT APPLICATION NUMBER: US/09/369,364A  
; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 1882  
; TYPE: PRT  
; ORGANISM: Homo sapiens ADAMTS-9  
; FEATURE:  
; NAME/KEY: MOD RES  
; LOCATION: (468)  
; OTHER INFORMATION: Xaa = C  
; NAME/KEY: MOD RES  
; LOCATION: (521)  
; OTHER INFORMATION: Xaa = Y  
US-09-369-364A-13  
Query Match 8.0%; Score 657; DB 4; Length 1882;  
Best Local Similarity 21.2%; Pred. No. 3.5e-44;  
Matches 288; Conservative 123; Mismatches 421; Indels 526; Gaps 70;  
QY 201 WSS---NSPCKTCSGSLPLGFRSRNKHMAIGGKCPLELLEKEACIVEGELL--- 254  
828 WNSHGPWQACSKPCQ-----GERKR-----KLVCTRESQTLTVS 861  
QY 255 -QQCPRYSWRTSEKCEQVSLLEQQDPHWHVT-----GPVCGGIGITREYVCAQSVFAA 308  
862 DQRCDELQPGHITPCGTGC-----DLRWHVASRSECSAQCGLGRTLDIYCAK----- 911  
QY 309 AALRAKEVSRPVEKALCVGPAPLPQLCNPTCDVCISWS--SANGLCIHENCHEPQK 366  
912 -YSRLDGKTEKVDGFC--SSHPKPSN--REKSGECNTGGRYSATWEC----- 956  
QY 367 KGFRTRQRHLMESTGPAGHCPHLVSVPCEDPMCYRWLASGICFPDHGKCGLGHRLK 426  
957 -----SKSCDGGTQRRR 968  
QY 427 AVCONDRGVEDVSGSLCPVPPPPPKKSC-EIPCRMDCVLSEWT--EWSQSCSNKNSDG 483  
969 AICVNTNRDVLDDSKCTHQEKVTIQRCEFP-----PQKSGDWSECLVTC-----G 1016  
QY 484 KQTRSRITLALAGE--GGKPCPPSQALQEHRLCNHDSCHQLHETSMPWPCSEDTLVT 540  
1017 KGHKHSQVWCQFGEDRLNDRMCDPETHKTSNQTQOPECAS--WQAGPWWCS----- 1067  
QY 541 LNAITGNGEATCGVIGITRRVFC-VKSHVGVMTKRCPSDSTRPETVRPCPLPKCKDC-- 597  
1068 -----VTCGGYQLRAVKCIIGTYSVVDNDCNATRTPTDQDCLEP---SCHP 1114  
QY 598 -----IVTAFSEWTPCPRMCAQNAVTQKSRYRIIIEAANGQCEPDT 641  
1115 PPAEPTRRSTYSAPRTQWRFGSWTPCSATCGK-----TRMYVSCRDENG-----S 1162  
QY 642 LYVEEREC-----EDVSLCPVYRWKPKWSPCILLVPSVWQGITGSSEACGKGLQTR 692

1163 VADESACATLPRPAKEBCSVTPCGMKALDWS--SVTCQGRATR 1208  
693 AVSCLSDNRAEMCECLKQINGPLLVQCTVPCREDCTTANSKFTPTSTNCEATKSR 752  
1209 QVMCVNYSDDHIDRSEC--DDYIPETDQCSM--SPC----- 1242  
753 RRLQTKSRKEKQDSDL--YPLVETELCPDSEFISQY--GNWSDCILPEG 801  
1243 -----PQTPDGLAQHPQNEIDYRPSASPSETHVLGNQWRTGPGAC----- 1287  
802 RREPHRGLRVAODSKECEGLRFRVACSDKNG-----RPVDPSCSSSGYQEX 851  
1288 -----SSTCAGSQRRVVVQDENGTYANDCVERIKPDEQRACESG----- 1328  
852 CVIPCPDCKLSDW--SWGSSSSCGIGVIRSRKWKKEKPYNGRCPKLDLKNVHEA 909  
1329 ---PCP-----QWYGNWGECKLGGGIRTR--LVVSORSNGERPP--DLSCIELDK 1374  
910 VPCYSECNOY-----SWVHEWSSCKINNELSLRSGGTSQSRKIRCVNTADGEGGAVD 963  
1375 PPDRQCNTHACPHDAWSTGFWSSC-----SVSCGRGHKQRNVYCMK--DGSHLE 1424  
964 SNLQNDQDEIPETQSC--SLMCPNECVMSW--GLWSKCPQSCDPTMORTRHLLRPSLN 1020  
1425 SYCCKHLAKPHGRKCRGRCP-----KWAGAWSCSVSNRGRVQORH--VGCQIG 1474  
1021 SRTCAEDSQVOPCLLNEN-----CFQFOYNTLWSTCOLSENAPCGQGVTRLLSCV 1072  
1475 THKIARETECPNTPRESECECQGRCPPLTWRAEWEQEC---TKTCGEGSRVKKVCV 1530  
1073 -----CSDGK--PYSMDQCEQHLEKPRQMSIPLCECVNVCOLSWTA--WTBC 1118  
1531 DNKNVHGAARDVSKRPYDRESCLO-----PC--EYV-----WTTGEWSEC 1571  
1119 SOTCGHGRMSRTRIIPTQEGRCPTELTOEKT--CP-----VTPCY--S 1162  
1572 SVTCGKYQKQLVSCSEIYTGKEN---YEYSQYTTINCPGTQPPSVHPVILRECPVSAT 1627  
1163 VVLGNWACKLEGCGGQGVQIRISUSCMVHSGSISHAA----- 1200  
1628 MRVGNWGSV---SCGVGMQSRVQCLTNEDQPSHLCHTLKPBERTKCRNVNVCNCELPQ 1684  
1201 -----GRVEDALGCE--MPFODSILKQCSVPCPDGCHLTWSEWSTCELTCDGR 1249  
1685 NCKEVKRLKAGED--GEYFLMIRGKLKIFCA-----GMHSDHPKEVYT--LVHGDSE 1734  
1250 SF-ETVGRQSRRTFIISFENQDSC-----PQVLET----- 1281  
1735 NFSEVVGHLNPTPCPNYNGSRDDCQCKRDYTAAGFSFQKIRIDLTSMQIITDLQFA 1794  
1282 -----RP---CTGGKCYHT-----WKASLWNN---NERTVNCORS-----D 1312  
1795 RTSEGHVPFATAGDCYSAKCPQGRFSINLXGTGLSLTESARWISQGNVAVSDIKKSPD 1854  
1313 GYNVTGGCSFQAPAAIROCIACRKPFSYCTQGGVCG 1350  
1855 GTRVVGKC-----GGYCG 1867

## :RESULT 3

IS-09-800-729-89

Sequence 89, Application US/09800729

Patent No. 6605592

## GENERAL INFORMATION:

APPLICANT: N. et al.

TITLE OF INVENTION: 32 Human secreted proteins

FILE REFERENCE: P204PL

CURRENT APPLICATION NUMBER: US/09/800,729

PRIOR FILING DATE: 2001-03-08

PRIOR APPLICATION NUMBER: PCT/US00/26013

PRIOR FILING DATE: 2000-09-22

PRIOR APPLICATION NUMBER: 60/155,709

PRIOR FILING DATE: 1999-09-24

NUMBER OF SEQ ID NOS: 217  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 89  
LENGTH: 1745  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-800-729-89

Query Match 5.8%; Score 477.5; DB 4; Length 1745;

Best Local Similarity 19.2%; Pred. No. 1.2e-29; Indels 499; Gaps 51;  
Matches 221; Conservative 124; Mismatches 318;

201 WSSWSPCSKTCRSG--SLLPGFRSRS---RNVKMAIGGKECEPEL---LKEACIVEG 251  
Db 39 WGSWSSCRSTCGGASYSLLRCLSSKSCBGRNIRYRTC--SNVDCPEAGDFRAQCSAHN 97  
Qy 252 ELLOQCPRYSWRITSEWKECOVSLLEQDPHHVITGVPVCGGIGIOTREYCAOSVPAARAL 311  
Db 98 DVKHGQFYEW-----LPVSNDP-----DNPCSLKC 123  
Qy 312 RAKEVSRPVEKALCVGPAPLPQCLNIPGSTDCTIVSSWSWAMGLCIHENCHBPQKKGFRT 371  
Db 124 QAKGTTLVVELA-----PKVLGTRCYTSSL-----DMCISGLC----- 157  
Qy 372 QRHVLMESTGPAGHCPHLVESVPCDDPMCYRWLASBGICFPDGHKCGLGHRLKAVCON 431  
Db 158 -----QIVG---CDHOLGSTVKED-----NCGVCNGDGGSTCLRVGGYQSLSA 198  
Qy 432 DRGEDV-----SGSLCPVPPPPPERKSCBIPCRMDCVLSEWTEWSSCSQSCSNKNSDGK 484  
Db 199 TKSDDTTVAIPYCSRHIRILVKGPDHLYLETK--TLOGTKGENSLSTGTFLVDNSSVDFO 257  
Qy 485 QTSRRTILALAGSGGKPCPPSOALQOHLRLCNHNSCNQL-----HWTSWGPCSEBT 536  
Db 258 KFPDKELRMAG---PLTADFIVKIRNSGSADSTVQFIYQPIIHRWRETDFPPCS--- 310  
Qy 537 LVLTALNATIGNGEANTCGVIGIOTRRVFCVKSHVQVMTKR-C---PDSRTPETVRCELP 592  
Db 311 -----ATCGGGYQLTSAECYDLRSNRVVADQYCHYYPENIKP---KPLQOE 353  
Qy 593 CKKDCIVTAFSEWTPCPRMCQAGNATVKOSRYRIIIEAANGQECQPDLYEBRECEDEVS 652  
Db 354 CNLD-----PCPA----- 361  
Qy 653 LCPVYWKQKWSFCILVPESVWQGITSGSEACGKLOTRAVSCISDDNR-----SAEMME 708  
Db 362 -----RWEATPWTAC-----SSCGGGIQRVAVSCVEEDIQGHVTSVEEWK 402  
Qy 709 CLKQTNMPLLVQECTVPCREDCTFTAWSKFTPCSTNCEATKSRRLQATCKSRKKEKQD 768  
Db 403 CM-YTPQPIA-----QPCN----- 416  
Qy 769 SDLYPLVETELCPDSEFISQYGNWSDCILPEGRPHRGLRVAODSKCEGGLRPRAVA 828  
Db 417 -----IFDCPKLAQ---EWSFCTV-----TCGQGLRYRVVL 445  
Qy 829 CSDKNGRPVDPSPCSSSGYTOEKVIP-----CPFDCKLSDW----- 865  
Db 446 CIDHRGHNHTGCGSPKTKPHIKECIVPTCYKPKELPVEAKL--PWFQAOQLEEGAAVS 504  
Qy 866 -----SSWGSSCSSCGIGVIR-----SKWLKEKPYN---GGRPCPKLDLKNQ 905  
Db 505 EEPFIFPKAWSACTVTCGVGTQVIRVCQVLLSFQSQVADLPIDECEGPKPA-----SOR 559  
Qy 906 VHEAPCYSECNQ-----YSWVVEHWSCKINNELSLRSLFCGGGTS 946  
Db 560 ACYAGCSGEIPFNDEDTGLFGGLQDFDELWDYEGFTK-----SSCGGQVQE 612  
Qy 947 RKIRCYNVADGEGGAVDSNLQNDQEIPE--TOSCSLMCFNECVMSWGLSKCPQSCDPH 1005  
Db 613 AVVSLNKKQTR--PAENLCVTSRRPPQLKSCNL-----DPCPA----- 651  
Qy 1006 TMQRTRHLLRPSLSNRSRTCAEDSQVOPCLLNENCFQYNTLWSTCOLSENAPCGQGV 1065

```
652 -----RWEIGKWSPCSILT-----CGVGLQ 670
1066 TRLLSC-----VCSGKVSNDQCEQHNLEKQPMSPICLVECVWNCQ 1108
671 TRDVFCSHLLSRMNETVILADELCRQKPESTVOACNR-----FNCP 712
1109 LSGWTA-WTECSOTCGH-----GRMSRTRFIIMPTQ--GEGRPCPTLTQKTCPC 1156
713 PAWYPAQWQPCSRTCGGVQKREVLCKQRMADGSFLELPETCSASKPACQACKDDCP 772
1157 VPCYVWLVGNWSACKLEGDCGSGVQIRSLSC--MVHSGSISHAAGRVEDALCEMPFQ 1214
773 -----SEWLLSDMTECST--SCGEGTQTRSACRKMKTG-----LSTVNVNSTLCPLPFS 821
1215 DSILKQLCSVPC 1226
822 SSIR-----PC 827

RESULT 4
US-09-132-769-3
; Sequence 3, Application US/09132769A
; Patent No. 6525023
; GENERAL INFORMATION:
; APPLICANT: Motoo Yamasaki
; APPLICANT: Kenji Shibata
; APPLICANT: No. 6525023uo Hanai
; APPLICANT: Akiko Furiya
; APPLICANT: Kaoru Miyamoto
; TITLE OF INVENTION: NOVEL VASCULAR SMOOTH MUSCLE CELL GROWTH FACTOR
; FILE REFERENCE: 11078
; CURRENT APPLICATION NUMBER: US/09/132,769A
; CURRENT FILING DATE: 1998-08-12
; EARLIER APPLICATION NUMBER: HE19-218491
; EARLIER FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 3
; TYPE: PRT
; ORGANISM: BOVINE
US-09-132-769-3

Query Match          5.3%; Score 440.5; DB 4; Length 807;
Best Local Similarity 21.2%; Pred No. 4.2e-27;
Matches 219; Conservative 123; Mismatches 340; Indels 351; Gaps 50;

QY 376 VLMESTGPA-----GHCPHLVESVPCDDPMCVRWLASGII-----410
DB 7 LLRLSRGPALLALALPLAVALAFSDETLDKVPKSEGVCSRIILAVQGTREGVTEFSLRVE 66
QY 411 CFPDHGKGLGHRIT-----LKAVCQNDRGEDVSGSLCFVPPPPERKSC 453
DB 67 GDPDFYKPGTSYRVTLSAAPPSPVFRGTLLAKENREGKEEDHAGT-----113
QY 454 EIPCRMDCVLSEWTEWSSCSQSNKNSDKQTRSRITILALAGEGKPCPPSQALQEHRL 513
DB 114 -----FOIIDEETQFWSNCPVAVTESTPRRTR-----142
QY 514 CNHSCMQLWETSPPWPCSEDILVLTALNATIGNGEATCGVIGIOTRRVFCVKSHVQVM 573
DB 143 -----IQVFWIAPPAG-----TGCVLKASI-----VQRRIIYF--QDEGSILT 178
QY 574 TKRC-PDSTRPETVRPCFLPKCKDCIVTAFSEWTPCPRMCQAGNATVKSRYRIIIEAA 632
DB 179 KKLCEQDSTFDGVTDKPILDC-----CACGTA-----KYRLTFY--G 213
QY 633 NGGCECDPTLYEECECDVSLCPYRWKPKNSPFCILVPESVWQGTGSEACG-----686
DB 214 NWSEKTHPKDYPR-----ANHWSAIIIGGSHSKNYVLYWEY 248
QY 687 KGLQTRAVSCISDDNRSANMECLKQTNMGPLLQVETVPCREDCTFTAWSKF--TPCSTN 745
```

```
DB 249 GGYASEGVKQVAELGSLSVKVEBIRQQSDBLTV-----IKAKAQPWAPLNVRAAPS 302
QY 746 CEATKSRRL-----TKSRKKEKCCODSDLYPLVETELCPCDFISQPYG 791
DB 303 AEFSDVTRHLMBSFLTMGSPSPDNVGLSABDICTKECGWVQKVQDLPDWA-----G 356
QY 792 NWSICILPEGRREPHRLRVQADSKCEGELRPRVACSDKNGRPVDPSPFCSSGYIQ--849
DB 357 TDS-----GVTVESPNKPIPOEKIRPLTSLD--HPQSPFYDPEGGSTQV 400
QY 850 -----EKC-VIP-----CPFDCKLSDWSWGSCSSS--CG 876
DB 401 ARVIERIARKEGQCNIVPDNDVIDVADLAFERKDEDDTPTCTIYSNWSPWACSSSTCD 460
QY 877 IGVRIRSKWLKEKPYNGRPRCPKLDLKNQVHEAVPCV-----SECNQYVWVVEHWS 927
DB 461 XGEMRQRMKQAQ-LDLSVFCP-----DTQDFQPCMGPCSDSDGSGTCTMSEWIT--WS 511
QY 928 SKINNELSLRCGGTQSRKIRCVNTADGEGAVDSNLNCNQEIP-PETQSCSL---MC 983
DB 512 PC-----SISCGTGRSRE-RYVKQFPEDG-----SVCT---LPTETEKTVAEES 555
QY 984 PNECVMSWGLWSKCPQSCDPHTMORRTRHLLRPSLNSRTCAEDSOVQPCLLNENCFQFO 1043
DB 556 PSSCLTTEWGEWDECSATCGMGMKXHRVMKSPADGSMCKAETSQAECMMEF-CHTIP 614
QY 1044 YNL---TENSTQLSENAPCGQVRT--RLLSVCSDGKPVSMDCQEQHNLKPKQMSIP 1098
DB 615 CLLSLWSEWSDCSVT---CGKGNRTQRMKSLAELG-----DCNE-ELEQVEKMLP 663
QY 1099 CLVECVNCSLGSWTATECSOTCGHGRMSRTRFIIMPTQSGRGPCTELTQKTCPTV 1158
DB 564 ---ECPIDCELTEWSQWSECKNSCGK-GEMIRMTQMEPQFGTFCP-ETVQRKKCRIR 718
QY 1159 PCYSWVLGNWSACKLEGDCGEGVQIRSLSCMVHSGSISHAAGRVEDALCEMPFQDSIL 1218
DB 719 KC---LRNPSIQNLWRARE-----SRSEQLREESDQ-----751
QY 1219 KQLCSVPCGDCHLTEWSEWSTCELTCDIG--RSFETVGRQSRRTFIQSFEHQDSCPQ 1276
DB 752 -----FPG-CRMRPMTAWSSECTKLGCGGIQBYRTVTKRFSOF-----TSCKD 795
QY 1277 QVLETRPCTGGKC 1289
DB 796 K-KEIRACNVHPC 807

RESULT 5
US-09-132-769-1
; Sequence 1, Application US/09132769A
; Patent No. 6525023
; GENERAL INFORMATION:
; APPLICANT: Motoo Yamasaki
; APPLICANT: Kenji Shibata
; APPLICANT: No. 6525023uo Hanai
; APPLICANT: Akiko Furiya
; APPLICANT: Kaoru Miyamoto
; TITLE OF INVENTION: NOVEL VASCULAR SMOOTH MUSCLE CELL GROWTH FACTOR
; FILE REFERENCE: 11078
; CURRENT APPLICATION NUMBER: US/09/132,769A
; CURRENT FILING DATE: 1998-08-12
; EARLIER APPLICATION NUMBER: HE19-218491
; EARLIER FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1
; LENGTH: 807
; TYPE: PRT
; ORGANISM: human
US-09-132-769-1

Query Match          5.3%; Score 434.5; DB 4; Length 807;
```

[illegible]

```

US-09-640-173-186
; Sequence 186, Application US/09640173
; Patent No. 6613515
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: OVARIAN TUMOR SEQUENCES AND
; TITLE OF INVENTION: METHODS OF USE THEREFOR
; FILE REFERENCE: 210121.484C2
; CURRENT APPLICATION NUMBER: US/09/640,173
; CURRENT FILING DATE: 2000-08-15
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 186
; LENGTH: 807
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-640-173-186

Query Match          5.3%; Score 434.5; DB 4; Length 807;
Best Local Similarity 21.1%; Pred. No. 1.3e-26;
Matches 216; Conservative 120; Mismatches 342; Indels 317; Gaps 50;

Qy      391 VESVPCEDDPMCYWLASEGI-----CFFDHGKCGLGHRLIKAVCQNDRGEDV 437
Db      34 LDKVPKSEGYCSRILRAQGRREGYTEFSLRVEGDDPYFKPGTSYRV-----80

Qy      438 SGLCPVPPPPPKRSCEIPCRCMDCVLSEWTWSSCSQSCSNKNSDGQTRSRITALLAGE 497
Db      81 --TLSAAPPSYFRGFTLIALR-----ENREGDKBEDHAGTFQIIDEE 120

Qy      498 GGR---PCPPSQALQEHRLCNDSHQMWHETSPWGPCSBDTLVTALNATIGNWGEATCG 554
Db      121 ETQFMSNCPVAVTESIPR---RRTRIQVFIAIPAG-----TGCVILKASI-----163

Qy      555 VGIQTRRVFCVSHVGQVMYKRC-POSTRETVPKPLPCPKDCIIVTAFSEWTPCPRMQ 613
Db      164 --VQKRIIYF--QDEGSLTKLCEQDSTFDGVTDKPIIDC-----CA 201

Qy      614 AGNATVKQSYRIIIQEAANGGCEPDTLYEEECEDVSLCPVYRWKPKQKSPCLLVPE 673
Db      202 CGTA-----KYRUTFY--GNWSKTHPKDYPRR-----AN 229

Qy      674 VWQGITGSSEACG-----KGLQTRAVSCISDDNRSAAEMCEIKQTNGMPLLVQECTVPC 727
Db      230 HWSAIIIGGSHSNVYLWEYGYASGKQVAELGSPVQVEEIRQOSDEVLTV-----1 283

Qy      728 REDCTITAMSKF-TPCSTNCEATKSRRLQ-----TGSKRKKKCEQDSLDY 772
Db      284 KAKAQPAWQPLNVRAPASAFBSVDTRHLSPLTWMGPSPDNWNVLASBDLCTKECGWV 343

Qy      773 PLVETELCPDCEFTSQPYGNWSCILPEGRREPHRGLRVQADSKCEGEGRLFRFAVACSDK 832
Db      344 QKVVDLLPWA-----GTDS-----GVYESPNKPTIPQEKIRPLTSLD- 383

Qy      833 NGRVPDPSPCSSSGYIQ-----EKC-VIP-----CPF 858
Db      384 --HPQSPFYDPEGSGITQVARVVIERRARKEQCNIQPNVDDIVADLAPEEKDEDDTPE 441

Qy      859 DCKLSDWSSNGSGCSSS-CGIGVIRIRSKWLKKEPYNGRCPCKLDLNQVHEAVPCY----913
Db      442 TCIYSNWSPWSACSSSTCDKGRMRBMLKAO-LDLISVFP-----DTQDFQPCMGPGC 494

Qy      914 -----SECNQYSVWBHWSCKINNELRLRCGGGTQSRKIRCVNTADGEGGAVDSNLCN 968
Db      495 SDEGSGTCTMSEWIT--WSPC-----SISCGMGWRSRE-RYVKOPPEDG-----SVCT 539

Qy      969 QDEIPE-----TQSCSLMCPNECVMSBWLKSKPCOSCDPHTMORRTRHLLRPSLNS 1021
Db      540 ---LPTHEMEKCTVNECS---PSSCLMTGWBDECSATCGMGKKRHRMIKNFADGS 593

Qy      1022 RTCAEDSQVQPCILLNENCFQYNL-----TEWS7CQLSENAPOCQGYRT--RLLSVCVCSDG 1076

```

594 MCAETSQAEBKMMPE-CHTIPCLLSPWSEWSDCSVT-----CGKGMRTQRMLKSLAELG 648  
1077 KPVSMQCEQHNLEKQRMSPICLVECVNQLSGWTAWTECSQTCGHGRMSRTRFIIM 1136  
649 -----DCNE-DLEQVEKMLP---ECPIDCELTEWSQSECNKSCGK-GHVIRTRMIQM 697  
1137 PTQGEGRPCTELTOEKTCPVTPCYSNVLGNWSACKLEGDCGEGVQIRSLSCWHSISI 1196  
698 EPQFGGAPCP-ETVQRKKCRK----LRNPSIQKPRWREARE----- 736  
1197 SHAAGRVEDALCCEMPQDSILKQLCSVPCGDCHLTEWSEWSTCELTICDG--RSFETV 1254  
737 SRRSEQLKEESEGEQ-----FPG-CRMRPMTAWSECTKLCGGGIQERYMTV 781  
1255 GQSRSRRTFIQSFENQDSCPOQVLETRPCTGGKC 1289  
782 KRFKSSQF-----TSCKDK-KEIRACNVHPC 807

RESULT 7  
US-09-713-550-186  
; Sequence 186, Application US/09713550  
; Patent No. 8617109  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
; THERAPY AND DIAGNOSIS OF OVARIAN CANCER  
; FILE REFERENCE: 210121.484C4  
; CURRENT APPLICATION NUMBER: US/09/713.550  
; CURRENT FILING DATE: 2000-11-14  
; NUMBER OF SEQ ID NOS: 205  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 186  
; LENGTH: 807  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-713-550-186

Query Match 5.3%; Score 434.5; DB 4; Length 807;  
Best Local Similarity 21.7%; Pred. No. 1.3e-26;  
Matches 216; Conservative 120; Mismatches 342; Indels 317; Gaps 50;

391 VESVPCDDPMCYRLASEGI-----CFPDHKGCGLGHRLKAVQNDRGEDV 437  
34 LDKVPKSEGVCSRLRAQGTREGEYTBFSLRVEGDDPFYKPGTSYRV----- 80  
438 SGSLCPVPPPPPERKSCIPCRMDCVLSEWTEWSSCSQSCSNKNSDGKQTRSRITILALAGE 497  
81 --TLASAAPPVYFRGFTLIALR-----ENREGDKEDHAGTFQIIDE 120  
498 GKG---PCPPSQALQEHRLCNDHSCMQLHWTSPWGPCSEDTLVTALNATIGNWGECATCG 554  
121 ETQFMSNCPVAVTESTPR---RRTRIQVFIAPAG---TGCVLKASI----- 163  
555 VGIOTRRVFCVKSHGVQVMTKRC-PDSTRPETHVPCFLPKCKDCIVTAFSEWTPCPRMCQ 613  
164 --VQKRIIYF--QDEGLTKLCEQDSTFDGVTDKPILDC-----CA 201  
614 AGNATVQSRVRIIIOBAANGQCEPPTLYEERECEDVSLCPVYRWKPKQKWSPIILVPS 673  
202 CGTA-----KYRLTFY--GNWSEKTHPKDYPR-----AN 229  
674 VWQIGTGSSEACG-----KGLQTRAVSCISDDNRSAMMECLKQTMGMLLVQECTVPC 727  
230 HWSAIIIGSHSKNVLYHEYGAYSEGKVAELGSPVWMEERINQGDVLTIV-----I 283  
728 REDCTFTAWSKF--TPCSTNCEATKSRRLQ-----TGKSRKXKCEQDSDLY 772  
284 KAKAQWPAWQPLNVARAAPSAEFSVDRTRHLSFLTMWGPSPDNVNVGLSAEDLCTKCGWV 343  
773 PLVETELCPCEDEFISQPGNWSDCILPEGRREPHRGLRVQADSKCEGGLRFAVACSDK 832

Db 344 QKVVDLIPWDA-----GTDS-----GVYESPNKPTIPQEKIRPLTSLD- 383  
QY 833 NGRPVDSPFCSSSGYIC-----EXC-VIP-----CPF 858  
Db 384 --HPOSPFYDPEGSSITQVARVIERIARKGEQCNIPVDNVDDIVADLAPEEKDDTPE 441  
QY 859 DCKLSDHSSWGSCESSS--CGIGVIRSRKWLAKPNGRCPCLDLKNQVHEAVPCY--- 913  
Db 442 TCIYSNPSWPSACSSSTCDKGRQRMLRAQ-LDLSVPCP-----DTQDFQCMGPGC 494  
QY 914 -----SECNQYSWVVEHWSCKINNELRLRCGGGTQSRKIRCVNTADGGGAVDSNLGN 968  
Db 495 SDEGSTCTMSEWIT--WSPC-----SISCGMGMSRE-RYVKQPPEDG-----SVCT 539  
QY 969 QDEIPPE-----TQSCSLMCPNECVMSWGLWSKCPQSCDPTMORRTEHLLRPLNS 1021  
Db 540 ---LPTMEMEKTVNEBES---PSSCLMTWGEWDECSATCGMGKMKRHRMKNVPADGS 593  
QY 1022 RTCAEDSQVQCLLNENCFOFYNL---TEWSTQLSENAPCGQVRT--RLISCVCSDG 1076  
Db 594 MKAETSQAEBKMMPE-CHTIPCLLSPWSEWSDCSVT-----CGKGMRTQRMLKSLAELG 648  
QY 1077 KPVSMQCEQHNLEKQRMSPICLVECVNQLSGWTAWTECSQTCGHGRMSRTRFIIM 1136  
Db 649 -----DCNE-DLEQVEKMLP---ECPIDCELTEWSQSECNKSCGK-GHVIRTRMIQM 697  
QY 1137 PTQGEGRPCTELTOEKTCPVTPCYSNVLGNWSACKLEGDCGEGVQIRSLSCWHSISI 1196  
Db 698 EPQFGGAPCP-ETVQRKKCRK----LRNPSIQKPRWREARE----- 736  
QY 1197 SHAAGRVEDALCCEMPQDSILKQLCSVPCGDCHLTEWSEWSTCELTICDG--RSFETV 1254  
Db 737 SRRSEQLKEESEGEQ-----FPG-CRMRPMTAWSECTKLCGGGIQERYMTV 781  
QY 1255 GQSRSRRTFIQSFENQDSCPOQVLETRPCTGGKC 1289  
Db 782 KRFKSSQF-----TSCKDK-KEIRACNVHPC 807

RESULT 8  
US-09-800-729-217  
; Sequence 217, Application US/09800729  
; Patent No. 6605592  
; GENERAL INFORMATION:  
; APPLICANT: Ni et al.  
; TITLE OF INVENTION: 32 Human secreted proteins  
; FILE REFERENCE: P2044P1  
; CURRENT APPLICATION NUMBER: US/09/800.729  
; CURRENT FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: PCT/US00/26013  
; PRIOR FILING DATE: 2000-09-22  
; PRIOR APPLICATION NUMBER: 60/155,709  
; PRIOR FILING DATE: 1999-09-24  
; NUMBER OF SEQ ID NOS: 217  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 217  
; LENGTH: 1059  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-800-729-217

Query Match 5.2%; Score 429; DB 4; Length 1059;  
Best Local Similarity 22.3%; Pred. No. 5.2e-26;  
Matches 263; Conservative 112; Mismatches 382; Indels 424; Gaps 73;

QY 418 CGLG-----HRLKAVQNDRGEDVSGSLCPVPPPERKSCIEIFCRMD-----CVLSWTE 468  
Db 39 CGGVSVRQLRCLTSKCS---GESVRFKVC-----AQKTCESKRLARDTICGEELVS 89  
QY 469 WSSCSQSCSNKNSDGKQTRSRITILALAG-----EGKPCPPSQALQEHRLCNDHSC-- 519  
Db 90 RQCEVVC-----RSR-----LTGANFLWRVDDGTPC---QAATSRVAVCSKSGCQI 132

520 -----MQLHWETSPWGPC-----SEDTLVTAALNA--TIGWNGEATCGV 555  
133 VCGDGLISSFFDAGVCGGSGDTCNKGFTIKWVSEETACASNCDDIVDWSAGRSIA 192  
556 GIQTRRVFCVSHVQVM-TKECPSTPET--VRPC-FLPCKKDCIVTAFSEWPCPRMC 612  
193 STSQIVVCNVAITGRVPEKLCADKLARKVAFRCPMLIIPSRWMA---ADWTECVPHC 249  
613 QAGNATVQSRVRIIIQEAANGQCEPDTLYE-----ERECEDVSLCPYRWKPKQWS 665  
250 --GEGTRKEV--CVQTAHVTVHVPDTFCENGTRPAAEENCVSTS-CG--RWEAGWS 302  
666 PCLIPVESVQGITSSSACGGLQTRAVSCISDDNRAEMECLOKOTNGMPLLVQECTV 725  
303 KC-----TASCGGVRRRHVAVGGS-----DC--DEGGRP----- 331  
726 PCREDCTFTAMSKFTPCSTNCEATKS-----RRRLTGKS-RKEXCQDSLDYPLVETELC 780  
332 --RQETTCVAG--IPCSI--ATNSLDNDRAYLDGNTFGSMNDHNDQAPRLVA---- 379  
781 PCDEFISQYGNWSDCILPEGRREPHRGLRVQADSKEGEGULRFRAVACSKNGRPVDP- 839  
380 -----GWSSTC-----SSTCGTGYMSRTVECVAVNPFISSAPI 411  
840 ----SFCSSSGV--IOEK-VIPCPF--DCKLSD-----WSSWGSSCSSCGIGVIR 882  
412 KLPMSBCQOEOPKIFESCEVRSCLQEDSKLSEDEAPYQWRYGDTQCSASC-LGGKQK 470  
883 S-----KW-----LKEKPYNGRGPCKLDLKNQVHEAVPCYSECNQVSWVVEH 925  
471 AALKCIQVSTGVSQVMSQCDARRRPPKSRPC-----NQ-HPCPP-----FWLTSK 515  
926 WSSCKINNELRLRCGGGTQSKIRCVNTADGEGAVDSNLNQNDE-----IPPETQSCS 980  
516 YSDC-----SMSGSGTARSVKAQTVSKTGA-DAHIVLRDRCHFKPKQBTETCN 567  
981 LM-CPNECVMS-----EWGLMSKQPCSDPHTMQR-----RTRHLRPP 1017  
568 VVACPATVSSLLKRNHKLKLNKLTAAQWTECSRSCDGERRRQVWCBIKRSRGTQRRP 627  
1018 SLNS-----RICAEDSQVQPCLLNENCFOQYNLTENSTCOLSENAPCGQVTRL 1068  
628 DVECDANTPQTVFVCSFGSGRPELLSNRVFE-QNABOKKLTGIGVATLYQTSIK- 685  
1069 LSCVC-----SDGKPYMSD-----QCEQHNLE--- 1090  
686 IKCPAKKFKKKIYKNGKKIKNDAAHKVANSANLNVPHARMEDAGVYECFTDLQGNV 745  
1091 -----KPO-----RMSIFCLVECVVNCOLSGWTAWTECSQTCGHGG 1126  
746 TLNFKYRDFPASRVDLAPKQIPSTKNRQVQVSKEDVLRQAS-----VLHKN 794  
1127 RMSRTRFIIMPQOGE-----RPPCTELTOEKTCPVTPCYSVNLGNWSACKLEGSDG-E 1180  
795 NVSLLEALLTADEXARQLRYKNELVAR-----WDIGHNSECROK--TCHVA 842  
1181 GVQIRSLSCMV-HSGSISHAAGVEDALCEMPFQDSILKQJCSV-----PC-PGDCHL 1232  
843 GYOARGISCKVTFHGBIRN-----VDNSIC-----ESLASVRPPETRPCHREDQFR 888  
1233 TEWSEWSTC-ELTCDIGRSFEIVGRQSRRTFIQSFENQDSCPOQ--VLETRPTCGKC 1289  
889 WBAQSWSECSORCVS-----SMLAQRRNVTRFTNGTSVDIOHCDITNRPATWDC 941  
1290 YHYT-----WKALWNNNRTVQCRSDGVNVTGG-----CSPQAR 1325  
942 PNQNCARWRTSDWGS-----CSSECG--TGGVQLRLLSCVWISSGRPAGENCEQMR 992  
1326 PAIROCI-----PACRKFPSVCTQ-GGVCGEKGYTEIMK 1360  
993 PHSARACVADEPLPPCMPTASALYQRDASQDSRPFCDIHK 1033

## RESULT 9

US-07-862-021B-10  
; Sequence 10, Application US/07862021B  
; Patent No. 5279966  
; GENERAL INFORMATION:  
; APPLICANT: Jessell, Thomas M  
; APPLICANT: Kiar, Avihu  
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A  
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooper & Dunham  
; STREET: 30 Rockefeller Plaza  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10112  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/862,021B  
; FILING DATE: 19920405  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: White, John P  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 40028  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 977-9550  
; TELEFAX: (212) 664-0525  
; TELEX: 422523 COOP UI  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 807 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-07-862-021B-10

Query Match 5.2%; Score 428.5; DB 1; Length 807;

Best Local Similarity 28.8%; Pred. No. 3.9e-26;

Matches 133; Conservative 64; Mismatches 158; Indels 107; Gaps 27;

Qy 849 QEKVTPCPDCKLSDWSWGSCTSS--CGIGVIRSKWLKEXPYNGRGPCKLDLKNQVH 907  
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 432 EEKEDDTPTCTIYSNWSPWSACSSTCEKGRMORMLKQ--LDLSVPCP-----DTQ 484  
Qy 908 EAVPCY-----SECQYSVVVEHWSCKINNELRLRCGGGTQSKIRKICVNTADGE 958  
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 485 DFQPCMGPGCDEDDGCTCTMSEWIT--WSPC-----SVSCGMGRSRE-RYVKQFPD 534  
Qy 959 GCAVDSNLCNQDIP-PEIQSCSL---MCPNECVMSWGLWSKCPQSCDPHTMQRTRHL 1014  
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 535 G-----SVC---MLPTEETKCTTNECSPSSCLVTEWGEWDDCSATCGMGMKRRMYK 586  
Qy 1015 LRPSLNRSTCAEDSQVQPCLLNENCFOQVNL---TEWSTCOLSENAPCGQVRT--RLI 1069  
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 587 MSPADGSMCKAETSQAEKCMPE-CHTIPCLLSPWSEWSDCVT---CGKGRTRQRL 641  
Qy 1070 SCVCSGDKPVSMDCQSOHNLKQPMRSIPCLVECVVNCOLSGWTAWTECSQTCGHGRMS 1129  
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 642 KSLAELG-----DCNE-DLEQAEKCMPL--ECPIDCELSWSEWSECNKSCGK-GHMI 690  
Qy 1130 RTRFLIPTQEGRPCTELTQETKCPVTPCYSVNLGNWSACKLEGDCGEGVQIRSLSC 1189  
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 691 RRTIQMEPFGGAPCP-ETVQKKCRARKC-----LRPSIQKLRWREARE----- 736  
Qy 1190 MWHSGSISHAAGVEDALCEMPFQDSILKQJCSVPCPGDCHLTEMSEWSTCELTICDG- 1248  
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 737 -----SRSEQLREESDGEQ-----FPG-CRMRPWTAWSECTKLCGGGI 774





CT-US93-03164-10

Sequence 10, Application PC/TUS9303164

GENERAL INFORMATION:

APPLICANT: Jessell, Thomas M

APPLICANT: Klar, Avihu

TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A

TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN

NUMBER OF SEQUENCES: 20

CORRESPONDENCE ADDRESS:

ADDRESSEE: Cooper &amp; Dunham

STREET: 30 Rockefeller Plaza

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10112

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US93/03164

FILING DATE: 19930402

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: White, John P

REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 40028

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 977-9550

TELEFAX: (212) 664-0525

TELEX: 422523 COOP UI

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 807 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: protein

CT-US93-03164-10

Query Match

Best Local Similarity 5.2%; Score 428.5; DB 5; Length 807;

Matches 133; Conservative 64; Mismatches 158; Indels 107; Gaps 27;

849 QEKVPCPFCDCKLSDWSWGSQSSS-CGIGVRIIRSKILKEPYNGGRPCPKLDLKNQVH 907  
432 EKDEDDTPTETCIYGNWSPWASCSSTCEKGRMRQRLKAO-LDLSPVPCP-----DTQ 484  
908 EAVPCY-----SECNQYGVVVEHWSCKINNELSLRCGGGTQGRKIRCVNTADGE 958  
485 DQPCMGPCSDDEGDTCTWSEWIT--WSPC-----SVSCGMGRSRE-RYVKQFPED 534  
959 GGAVDNLCNQDEIP-PETQSSCL---MCPNVCWSEGLWSKCPQSCDPHTMQRTTHL 1014  
535 G-----SVC---MLPTEETKCTVNEECSPSSCLVTWGEWDDCSATCGMGKKGRHVMK 586  
1015 LRPLNSRTRCARDQVOPCLLNNCFQFYNL---TEWSTCOLSENAPCGQGVRT--RL 1069  
587 MFADGSMCKATSKQAECWMPF-CHTIPCLLSPWSEWSDCVT---CGKGRTRFQRL 641  
1070 SCVSDGKPVSDQEQHNLKRPQMSIPCLVECVNCOLSGWTATWTCSPQCHGGRMS 1129  
642 KSLAELG-----DCNE-DLEQAEKMLP---ECPIDCELSWSQWSECNKSCGR-GHMI 690  
1130 RTFLIMPTQSGRCPPTLTQKTCPTVPCYSWLGNWSACKLEGGDCGEGVQIRSLSC 1189  
691 RRTIQMEPQFGAPCP-STVORKKCRARKC---LRSPSIQKLWREARE----- 736  
1190 MVHSGSISHAAGRVEDALCGEPFQDSILKQLCSVPCGDCHLTWSEWSTCELTCIDG- 1248

Db 737 -----SRRSEQLREESDGEQ-----FPG-CRMRPWTAWSECTKLCGGGI 774

QY 1249 -RSFETVGRQSRRTFIIOFENQDSCPOQVLETRPCTGKGC 1289

Db 775 QERYMTVKRKFSSQF-----TSCKDK-KEIRACNVHPC 807

RESULT 13

US-07-862-021B-12

Sequence 12, Application US/07862021B

Patent No. 5279966

GENERAL INFORMATION:

APPLICANT: Jessell, Thomas M

APPLICANT: Klar, Avihu

TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A

TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN

NUMBER OF SEQUENCES: 20

CORRESPONDENCE ADDRESS:

ADDRESSEE: Cooper &amp; Dunham

STREET: 30 Rockefeller Plaza

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10112

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/862,021B

FILING DATE: 19920405

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: White, John P

REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 40028

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 977-9550

TELEFAX: (212) 664-0525

TELEX: 422523 COOP UI

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 802 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: protein

US-07-862-021B-12

Query Match 5.1%; Score 419; DB 1; Length 802;

Best Local Similarity 23.8%; Pred. No. 2.3e-25;

Matches 186; Conservative 100; Mismatches 277; Indels 220; Gaps 44;

605 WT--PCPRMCOAGNATVKOSRY-----RITIOBAANGQECBPTLYEERECEV 651  
142 WTAPPTGTGCVILKASIVQRIYIFQDEGSLTKRICEQDSASEGVTDKPTL-----DC 194  
652 SLCPV--YR-----WKPQKMSPCILVPSVWQGITGSSEACG-----KGLQTRAVSCI 697  
195 CACGTAKYRLTFYGNWSEKTHPKDFPRRTNHWSAIIIGSSHKNYILWEYGVYASEGVKQV 254  
698 SDDNRSNEMECILKQINGMPLLVQECTVPCREDCTTANSKF-TPCSTNCEATKSRRL 756  
255 AELGSPVYMEERIKQSDSEVLTV-----IKAKAQWPAWQPLNVRAPSAEFSVDRHRL 308  
757 -----TGKSRKKEKQSDSLYPLVETELCPDCEFIISQPYGNWSDCILPEGR 802  
309 MSFLTMLGPSDNVGLSABDLCTKCGWQVQVQDLIPWA-----GTDS----- 354  
803 REPRHGLRVQADSKCEGEGRLFRVACSKDNGRPVDPSPCSSSGYIQ----- 849  
355 -----GVTYSPNKPVTPOKIRFLTSLD---HPQSPFYDPEGGSIKLVARVVLRIARK 406



TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham  
STREET: 30 Rockefeller Plaza  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10112  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/03164  
FILING DATE: 19930402  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 40028  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 977-9550  
TELEFAX: (212) 664-0525  
TELEX: 422523 COOP UI  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 802 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
CT-US93-03164-12

Query Match 5.1%; Score 419; DB 5; Length 802;  
Best Local Similarity 23.8%; Pred. No. 2.3e-25;  
Matches 186; Conservative 100; Mismatches 277; Indels 220; Gaps 44;

17	605	WT--PCPRNCQAGNATVKQSRV-----RIIQEAANGQCEPDILYEERECEHDV	651
18	142	WTAPPTGTCVILKASIVQKRIIYPODEGSLTKRICEQDSASEGVTDKPTL-----DC	194
19	652	SLCPV--YR-----WKPKWSPCLVPESVWQGITGSEACG-----KGLQTRAVSCI	697
20	195	CACGTAKYRLTFYGNWSEKTHPKDPRRTNHSALIGSSHSHKYNILWEYGVYASEGVKQV	254
21	698	SDNRSAEWMECLKQTNMPLLVQECTVPCREDCTFTAWSKF-TPCSTNCEATKSRRLQL	756
22	255	AEIGSPVMEEEIRQOSDEVLTV-----IKAKAQPAPQPLNVRAAPSAEFSVDRHRL	308
23	757	-----TGKSRKKEKQDSLDLYPLVETELCPCEBFISQPYGNWSDCILPEGR	802
24	309	MSFLTMLGSPDNVWGLSAEDLCTKCGWQVKVQDLIPWA-----GTDS-----	354
25	803	REPHRLGVQADSEKGEGLRFRVAVACSKNKRPNVDPSPFCSSSGYIQ-----	849
26	355	-----GVTVESPNKPTVPOEKIRPLTSLD---HPQSPFYDPEGGSIKLVARVVLRIARK	406
27	850	-EKC-VIP-----CPPDCKLSDWSSWGSCSSS-CGIGVIRSRKMLK	887
28	407	GEQCNFVDNIDIVADLAPEEKEEDDTTETCIYSNWSPWASACSSSTCEKGRMQRMLK	466
29	888	EXPYNGRPPCKLDLKNQVHEAVPCY-----SECNQYKVVVHSSCKINNLRSL	938
30	467	AQ-LDLSPVCF-----DTQDFQCMGPGCSDBDGSTCMWSDWIT--WSPC-----SV	510
31	939	RCGGGTQSKIRICVNTADGGGAVDNLQNDIIP-PETQSCSL---MCPNECVMSWGL	994
32	511	SCGMGTRSRRE-RYVKOFFPDG-----SMC---KVPTEETEKCIVNEECSPSSCLVTWGE	561

Qy	995	WSKCPOSQDPHTMQRTRHL-LRPSLNSRTCAEDSQVQPCLLNENCFQFYNL---TEWS	1050
Db	562	WDECSASCGT-GMKRRHRMTKMTPADGSMCKAETTEAEKMMPE-CHTIPCLLSPWSEWS	619
Qy	1051	TCOLSENAPCGQGVRT--RLLSCVCSGDKPVSMDQCEQHNLKFPQRMSPCLVECVVNCQ	1108
Db	620	DCSVT---CGKGMRTQRWLKSAELG-----DCNE-ELEQAEKMLP---ECPIDCE	665
Qy	1109	LSGWTAWTECSQTCGHGGRMSRTRFIIMPTQGEGRPCPTTELTOEKTCPVTPCVSWVLGNW	1168
Db	666	LTEWSQWSECNTSCGK-GHMIRTRMIKIEPPQFGGTACP-ETVQRTKCRVRKCLR-----	717
Qy	1169	SACKLEGGDCGEGVQIRSLSCMVHSGSISHAAAGRVEDALCGEMPFDOSILKQLCSVPCPG	1228
Db	718	-----GPGWEKR-----RWKEA--REKRSSEQAKQNDNEQYP-748	
Qy	1229	DCHLTWSEWSTCELTICIDG--RSFTVGRGSRSTRFIIQSPFNQDSQPQVLETRPCTG	1286
Db	749	VCLRKPTAWTEGTLCCGGIQERYVMVKRSKSTQF-----TSCKDK-KELRACNV	799
Qy	1287	GKC	1289
Db	800	HPC	802

Search completed: February 24, 2004, 01:09:33  
Job time : 64.7866 secs

4296

```
1223 SVPCGCHLTENSEWSTCBLTCTIDGRSFTVGRQSRSTFIITQSPENQDSQOVLTR 1282
1239 SPCGCGCYLKWSSWSLQLTCTVNGEDLGGGIGVRSRFTIITQELNQHLCPEQMLETK 1398
1283 PCTGGKCVHTWKASLNNNERTVWCORSQDVNTGCSPOARPAALROCIIPACRKPFSY 1342
1399 SCYDGCYCYEYKXWASANKGSRRTVWCORSQDVNTGCSPOARPAALROCIIPACRKPFSY 1458
1343 CTQGGVCGCEKGYTEIMKSNGLDYC-----MKVPGSEDKKADVKNLSGKRRPVNSKIH 1397
1459 CSETKTCHCEGYTEVMSSNSTLEQCTLIPIVWLPTMEDKRGDVK-SRAVHPTQPSNP 1517
1398 IPKG--KSLQPLDGDGKLVVYGVSGAFILMTFLFTSYLVCKKP-KPHQSTPPQOKP 1454
1518 AGRTWFLQPGFDGRLKTVIGVAGAFVLLIFIVSMYILACKKPKKQRRNNRLKP 1577
1455 LTLAYDGDLD 1465
1578 LTLAYDGDAD 1588

RESULT 2
US-10-074-566-2
; Sequence 2, Application US/10074566
; Publication No. US20030207348A1
; GENERAL INFORMATION:
; APPLICANT: Shinkets, Richard A.
; APPLICANT: Fernandes, Elma R.
; APPLICANT: Li, Li
; APPLICANT: Gorman, Linda
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Patturajan, Meera
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Spytek, Kimberly A.
; TITLE OF INVENTION: Polypeptides and Polynucleotides Encoding Same
; FILE REFERENCE: 15966-556 CIP1
; CURRENT APPLICATION NUMBER: US/10/074,566
; CURRENT FILING DATE: 2002-02-13
; PRIOR APPLICATION NUMBER: 09/619,252
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 60/144,722
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: 60/167,785
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: 60/276,994
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/280,898
; PRIOR FILING DATE: 2001-04-02
; PRIOR APPLICATION NUMBER: 60/332,241
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: 60/288,062
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: 60/291,766
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 60/314,007
; PRIOR FILING DATE: 2001-08-21
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1588
; TYPE: PRT
; ORGANISM: human
US-10-074-566-2

Query Match 52.1%; Score 4296; DB 15; Length 1588;
Best Local Similarity 50.7%; Pred. No. 0;
Matches 766; Conservative 238; Mismatches 449; Indels 58; Gaps 20;

2 VECIOKLNRTVTVANICHEALQPTQACILPCRDVCVSEFLPWSNCSKCGKLOHR 61
89 IACIQKNDIPAEIDICIFEFKPLLEQALIPCCQDCIVSEFSAWSECSKTCGLOHR 148
```



Y 62 TRAVIAPPLFGGLQCPNLTESRACDAPISCPGLBEEYFTSLKVGPMWCKRLPHLKEINPS 121  
Y 149 TRHVAVPQFGSGCPLNTEFQVQ---SSPCEAEELRYSLHVGPMWTCSPHSPQRQA 205  
Y 122 CRT-----VLDFNSDNERVTFK-HOSYKAHHKSKWAIEIGYQTRQNSCTR 167  
Y 206 RRRGNKEREKDRSGVXD--PEARELIKRRNRNRQNRQENKYWDIQIGYQTRVEMCIN 263  
Y 168 SDGQVAMLSLCLQDSFPLTVQSCIMPDCETSSQWSSPSCSKTCSGSLLPFGFRSRNV 227  
Y 264 KTGRAADLSFCQEKLPMTFQSCVITKECVSESENSEWSPCKTCHDMVSPAGTAVRTTI 323  
Y 228 KMAIAGGKCEPPELLEKACIVBELLOQCPRYKWRSEWKECOVSLLEQQDPHHWHTG 287  
Y 324 RQFPFGSEKCEPEFEKPCISQSDGVVFCATYGNRTTEWTECEKRVDPILLSQDXXRGNQT 383  
Y 288 PFCGGGOTREYVCAQS---VPAALAAKAEVRPEKALCVGPAPLPSQLCNIPGSTD 343  
Y 384 ALCGGGIOTREYVCVQANENLLSQLSTHKNKASKPMDLKLCTGPIFNTTQLCHIPTE 443  
Y 344 CIVSWSAWGLCIHENCHBPQKGRFTRORHVLWSTGPA---GHCPHLVESVPCEDPM 400  
Y 444 CEVSPWSAWGPTCYENCNDQCKGKFKLRKRITNEPTGGSGVGTGNCPLLEAIPCCEPA 503  
Y 401 CYRWLASE-GICFPDHGK-CGLGHRILKAVQNDGEGDVSGSLC-----PVPPPERKSC 453  
Y 504 CYDKAVRLGDCBPNGKECGPGTQVQEVVINCINGDEEVDRLQCRDAIFPFP-----VAC 558  
Y 454 EIPCMDCVLSEWTEWSSCOSCSKNKSDGKQTRSRITLALAG-EGGKPCPPSQALQEHR 512  
Y 559 DAPCPKCVLSTWSTWSSCHTSCSKTTEGQIRARSILAYAGEGGIRCPNSALQEV 618  
Y 513 LNDHSCMLHWEISWPWPCSEDTLVTALNATIGNGEATCGVGIOTRVRFCVKSHVQOV 572  
Y 619 SCNEHPCVYHWQTPWQCIEDTSVSFNTTTWNGEASCSCVGMQTRKVICRVANVQOV 678  
Y 573 MTKRCPDSTRBETVAPCFPLPCKCKCIVTAFSEWTPCPRYMCOAGNATV-KQSRYYRIIOEA 631  
Y 679 GPKKCPESLRBETVAPCLLPCKCKCIVTAFSEWTPCPRYMCOAGNATV-KQSRYYRIIOEA 738  
Y 632 ANGGQECPTLYEERCEBDSVLCPIYRWKPKQWSPCILIVPESVWQINGTSSEACGKGLQT 691  
Y 739 ANGGRDCTDPLYEEKACEAPACQSYRNKTHKWRRCQLVPVWSQDSGPAQEGCGPQQA 798  
Y 692 RAVSCISDNRSAENMECLKQTNMPLAVQECTVPCREDCTFTAKSKFTPCSTCEATKS 751  
Y 799 RAITCRKQGGQAGHECLQVAGVPALTAQCIQPCDDCLTNSKFSNCGSGAGVRT 858  
Y 752 RARQJTKSRKKEKQSDLYPLVETELCPDDEFISQPYGNMSDCILPEGRREPHRGLRV 811  
Y 859 RKRTLVGSKKKECKNSHLYELIETQYCPCKYNAQPVGNWSDCILPEGKVEVLGMKV 918  
Y 812 QADSKECEGLFRFRAVACDKNGRPVDFSCSSGYIOEKVIPCDFPCKLSDRWSWGC 871  
Y 919 QGDIKEGCGYRYQAWCYDQNGRLVETSRNCHGYIIEACIIPCSDCKLSEWNSRC 978  
Y 872 SSSCGIGVIRSKWLKEKPYNGRCPKLDLKN--QVHEAVPCYSECNQYSWVVERWSSC 929  
Y 979 SKSGGVKVRKWLREKPYNGRCPKLDHVNQVYVWVPCVCHDCNQYLWVTEPWSIC 1038  
Y 930 KINNELSRUCGGGTQSRKIRCV-NTAGEGAVDSNLGNODEIPPTQSCSLMCPNECV 988  
Y 1039 KYTFVNMRENCEGVQTRKVRQCMQNTADGSEHVEDYLCDPEEMPLGSRVCKLPCPEDCV 1098  
Y 989 MSEGWLKSCQSCDPHTMORTHLR--PSLNSRTCAEDSQVQCLNENCFQFYNLT 1047  
Y 1099 ISEWGPWTQCVLPCHQSSFRQASDPIQPADEGSCFNAVEKEFCNLNKNKYHYDNYT 1158  
Y 1048 EWSCTQLSNAPCGQGVTRLLSCVCSGDKGPKVSMQCEQHNLEKQPMRSPICLVECVVNC 1107  
Y 1159 DWSTQLSKAVCGNGIKTRMLDKVRSBGSKVDLKYCEALGLEKKNQWNTSMQVECPVNC 1218  
Y 1108 QLSGWTAWTECSQTCHGSGRMSTRFPIIMPTQGEGRPCPTELTQKTCPTVPCYSWVLN 1167

DB 1219 QLSDWSPWSECSTCGLTGKMLRRRTVTQPFQDGRPCPSPSLMDQSKPCPVKPCYRWQYQ 1278  
QY 1168 WSAKLEGDCGEGVQIRLSLSCWZHSGLSHAAAGVEDALCGEMPF-----QDSILKQLC 1222  
DB 1279 WSPQVQEAQOCGEGRTRNISCWSDGSDADDKSKVDDEFCADIELIIDGNKNMVBESC 1338  
QY 1223 SVPCPDCHLTSEWSTCELTICIDGRSPFETVGRQSRSTFTIIQSFENQDSCPOQVLETR 1282  
DB 1339 SQPCPDCLYKDWSSNSLCQLTCVNGEDLFGGIGVRSRPVIIQELNCHLCPEQWLETK 1398  
QY 1283 PCTGSKCYHTWKASLWNNERTVWCQSDGVNVTGGSQPARPAIROCIACRKPSPY 1342  
DB 1399 SCYDGCYCYKMAWAKWSSRTVWCQSDGVNVTGGSQPARPAIROCIACRKPSPY 1458  
QY 1343 CTQGGVCGCEKGYTEIMKNGFLDYC-----MKVPGSEDKADVKNLSGKPEPVNSKHD 1397  
DB 1459 CSEKTKCHEEGYTEWSSNSTLEQCTLIPVVVLPMTMEDKRGDKVT-SRAVHTQPSNNP 1517  
QY 1398 IFKG--WSIQPDDPGRVKIWIYGVGGAFLLIMIFLFTSYLVCKKP-KPHOSTPPOQKP 1454  
DB 1518 AGRGRTRWLPQFGPDGRKLTWYGVGAAGFVLLIFIVSMYILACKPKKPPQRRNNRKP 1577  
QY 1455 LTLAYDGDLDL 1465  
DB 1578 LTLAYDGDADM 1588

## RESULT 3

US-10-074-566-41  
; Sequence 41, Application US/10074566  
; Publication No. US20030207348A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimkets, Richard A.  
; APPLICANT: Fernandes, Elma R.  
; APPLICANT: Li, Li  
; APPLICANT: Gorman, Linda  
; APPLICANT: Gusev, Vladimir Y.  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Shenoy, Suresh G.  
; APPLICANT: Spytek, Kimberly A.  
; TITLE OF INVENTION: Polypeptides and Polynucleotides Encoding Same  
; FILE REFERENCE: 15966-556 CIP1  
; CURRENT APPLICATION NUMBER: US/10/074,566  
; CURRENT FILING DATE: 2002-02-13  
; PRIOR APPLICATION NUMBER: 09/619,252  
; PRIOR FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 60/144,722  
; PRIOR FILING DATE: 1999-07-20  
; PRIOR APPLICATION NUMBER: 60/167,785  
; PRIOR FILING DATE: 1999-11-29  
; PRIOR APPLICATION NUMBER: 60/276,994  
; PRIOR FILING DATE: 2001-03-19  
; PRIOR APPLICATION NUMBER: 60/280,898  
; PRIOR FILING DATE: 2001-04-02  
; PRIOR APPLICATION NUMBER: 60/332,241  
; PRIOR FILING DATE: 2001-11-14  
; PRIOR APPLICATION NUMBER: 60/288,062  
; PRIOR FILING DATE: 2001-05-02  
; PRIOR APPLICATION NUMBER: 60/291,766  
; PRIOR FILING DATE: 2001-05-17  
; PRIOR APPLICATION NUMBER: 60/314,007  
; PRIOR FILING DATE: 2001-08-21  
; NUMBER OF SEQ ID NOS: 132  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 41  
; LENGTH: 1588  
; TYPE: PRT  
; ORGANISM: human  
US-10-074-566-41

Query Match 52.1%; Score 4296; DB 15; Length 1588;



```
333 KTGKAADLSFCQOEKLPMTFQSCVITKEQVSEWSWSCSTKCHDMVSPAGTRVTRTTI 392
228 KMAIGGGKECEBELLKEACIVEGELLQCPRYSWRTSEWKEQVSLLEQDDPHWHTG 287
393 RQPIGSEKECEFEKEPCLSQGDGWPCATYGTWTEWTECRVDPPLLSQQDKRGHQT 452
288 PVGCGGIQIREVYCAQS---VPAALAKAVSRVEKALCVGAPLPSOLCNIPTCSTD 343
453 ALCGGIGIQIREVYCVQANENLSQLSTHKKAKSFMDLKLCTGPIPTNTQCHIPCPT 512
344 CIVSSWSAGLCHENCHPEPQGGKGRTRQRHVLKSTGPA---GHCPLHVESVPCEDPM 400
513 CEVSPMSAWGPTCYENCNDQGGKGFLLKRITNEPTGGSGVTGNCPELLEAIFCEBPA 572
401 CYEWLASE-GICPPDHGK-CGLGHRILKXAVCONDRGDEVSSGLC-----PVPPPERKSC 453
573 CUDKAVRUGNCEPDNGKSCGPGTQVQVVCINSQGEVDRQLCDAIFPIP-----VAC 627
454 EIPCRMDCVLSEWTESSQSCSNKNSDKGTRSRITLALAG-EGGKPCPPSQALQHR 512
628 DAPCPKDCVLSWTSSCSHTCSGKTTEGKQIRARSILAYAGEGGIRCPNSSALQVR 687
513 LCNHSCMOLHETSPWGCSEDTLTALNATIGNGENTCGVIGIOTRRVFCVKSHGVQV 572
688 SCNEHPCTVYHQTGWGQCIEDTSVSSFNFTTWNGEASCSVGQTRKVICRVNVGQV 747
573 MTKRCPDSTRPVRPCFLPKCKDCIVTAFSEWTPCPRMCOAGNATV-KOSRYRIIQEA 631
748 GPKCPESELRPETVRCLLPCKKDCIVTSPYDWTSCPSCKEGDSIRKQSHRVIIQLP 807
632 ANGQBCPTLTYEERCEVDVSLCPVYRWKPKQKSPCLVPESVWGQITGSSEACGKGLQT 691
808 ANGGRDCTDPLYEKACEAPACQSQSRWKTTHKWRRCQLVPMVSQDQSPGAQEGCGPGRQA 867
692 RAVSCISDDNRSAEMWECILQTNQWPLVQECTVPCREDCTFTAKSKFTPCSTNCEAKTS 751
868 RAITCKQDGGQAGIHECLQYAGVPALTOAQIPCDQDCLTSSKSSNCGDCGAVRT 927
752 RRQL 756
928 RRRL 932

RESULT 5
US-09-864-761-43221
Sequence 43221, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aeonica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
```

```

PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
SEQ ID NO 43221
LENGTH: 247
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC011231.1
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1
OTHER INFORMATION: SWISSPROT HIT: P35446, EVALUE 5.00e-10
OTHER INFORMATION: EST_HUMAN HIT: AW612526.1, EVALUE 3.00e-27
US-09-864-761-43221

Query Match 10.3%; Score 852; DB 9; Length 247;
Best Local Similarity 100.0%; Pred. No. 1.1e-58;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVRCIQKLNRTVVANEICHFALQPTTEQACLIQPCRDVCVSEFLPWSNCSKCGKKLQH 60
DB 95 MVRCIQKLNRTVVANEICHFALQPTTEQACLIQPCRDVCVSEFLPWSNCSKCGKKLQH 154
QY 61 RTRAVIAPPLFGGLQCPNLTESRACDAPISCPGEBEYTFSLKVGWPSKCRPLHLKEINP 120
DB 155 RTRAVIAPPLFGGLQCPNLTESRACDAPISCPGEBEYTFSLKVGWPSKCRPLHLKEINP 214
QY 121 SGRVLDFNSDSNERVTFKHQSYPKHHHSKSWA 153
DB 215 SGRVLDFNSDSNERVTFKHQSYPKHHHSKSWA 247

RESULT 6
US-09-321-987B-2
Sequence 2, Application US/09321987B
Patent No. US20020102210A1
GENERAL INFORMATION:
APPLICANT: Kimble, Judith E
APPLICANT: Brelloch, Robert H
TITLE OF INVENTION: Agent and Method for Modulating Cell Migration
FILE REFERENCE: 960296.95386
CURRENT APPLICATION NUMBER: US/09/321,987B
CURRENT FILING DATE: 1999-05-28
PRIOR APPLICATION NUMBER: 60/087,170
PRIOR FILING DATE: 1998-05-29
PRIOR APPLICATION NUMBER: 60/129,023
PRIOR FILING DATE: 1999-04-13
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 2150
TYPE: PRT
ORGANISM: Caenorhabditis elegans
US-09-321-987B-2
```

Query Match 8.4%; Score 694; DB 9; Length 2150;  
Best Local Similarity 22.9%; Pred. No. 4,9e-45;  
Matches 351; Conservative 133; Mismatches 556; Indels 492; Gaps 84;

QY 18 CEHFALQPTQACILPCPR-----DCVSEFLPWSNCKGCKKQHRTRAVIAP-PLF 71  
DB 571 CHH-----GACVRLAPESLTKIDQGWDRWSGECSTRCTGGVQKGLRDCDSPKRN 622

QY 72 GGLQCPNLTES-RACDAPISCPLEEYTFSLKVGPMWKRLPHLKEINPSGRTVLDPNS 130  
DB 623 GGYKCVGQGRYRSCNTQ-ECFMDTQPY-----REVQCS-----EFN- 658

QY 131 DSNERTFKHOSYAHHSKSWAIEIGYQTRQVSCSTRSDGQNALSL-----CLQD 181  
DB 559 --NKDIGIQGVASTNTHWPRIYANVNERCKLYC-RLSGSAFYLLRDKVDTGTPCDRN 715

QY 182 SFPLTVOSCMKDC-----TSQWSSWSPCKTCRSGSLLP-----219  
DB 716 GDDICVAGACMPAGCDHGLHSTLRDXGCVGGDSCKVKVGTFFNQGTFGYNEVMKIP 775

QY 220 FRSSRNKVMHAIIGGKECPLEKEACIVEGELLQOCPRYSWRTSEWKECQVSLLEQ- 278  
DB 776 AGSANIDIRQKGYNNKEDDNYLSRA--ANGEFLN-----GHFQVSLARQOI 822

QY 279 --QD-----PHWHVTGP-----VCGGIGTREV---YCAOSVPAALRA 313  
DB 823 AFQDTVLEYSGDALIERINGTGIRSDIYVHVLVSGSHPPDISYETWTAAPNA----- 877

QY 314 KEVSRPVEKAL-----CVGPAP-----LPSQ- 334  
DB 878 --V-RPISSALYLWRVTDWTECDRACRQGSQKLMCLDMSTRHQSHDRNCONVLKPKQA 935

QY 335 --LCNIPGSTCIVSSGAWGLCTHENCHEPQKGFRTQRHVLMESTGPAGCHLVE 392  
DB 936 TRMCNIDCSTWITEDVS-----SCSAKCGSGQRQRVSCVKMGEGDQRTFASHELCD 987

QY 393 --SVPCDEPMCY-----RWLASGI-CFPHGKCGLGRILKAVCONDRGEDVSGSLC- 442  
DB 988 RNSKPSDIASCYIDCSGRKNMYGWTSCSETCGSNGKMR--KSYCVDDSNRRVDESLOG 1045

QY 443 -PVPPPPERKSCIEPCRMDCVLSWT--EWSSCSOSCSNKNKSDGKQTRSTILLAGE-- 497  
DB 1046 REQKATERECNRIPC-----PRWTVGHSECSRSCDG---GVKWRHAQCLDAADRET 1095

QY 498 GKGKPPPSQALQEHRLCNHDSMOLHWETSPWGPCSEDTLVLTALNATIGNWGEATCGVGI 557  
DB 1096 HTSCGGAQT-QEH--CNEHAC--TWQFGVWSDCS-----AKCGDGV 1133

QY 558 QTRRVFCVKSIVGVQMTKCPDSTRPTETVRPCFLPKKDCIVTAFSWTFCPRMCQ----- 613  
DB 1134 QYRDANCTDRHSVLPBHRCLMKMKIIT-KPCH---RESCPKYKLGWSQCSVSCEDGWS 1189

QY 614 -----AGNAT-----VKOSRYRIIIEAANGQBCP-----639  
DB 1190 SRRVSCVSGNGTEVDMSLCGTASDRPASHQTCNLG-TCPPWRNTDMSACSVSCGIGHRR 1248

QY 640 --DTLYERECEDESVC-----PVYRWKPKWSPCILVPESVWQGITGSS 682  
DB 1249 TTECIYREQSV-DASFCGDTKWPETSQTCHELLPCTSWKPSHWSPC-----S 1293

QY 683 EACGKGLQTRAVSC-----ISDONSRAEMWE-CLKQTNMGMLPLAVQSCVPCRE 729  
DB 1294 VTCSGIGITRVSCTRSEGTIVDEYFCDRNTRPLKTKTEKTDGCPRVLQK-----LQA 1349

QY 730 DCTFTAMS--KFTPCSTNCEATKSRRLQTKSRKKEKCD-----SDLYPLVET 777  
DB 1350 DVPPRMTATGFWTACATCG-----NGTQBRLLKCRDHVRDLDPDECNHLKREVS 1400

QY 778 ELCPDDEFISQPYGNWSDCILPEGRREPHRLRQVADSKEGEGLEPRVACSKDKGRPV 837  
DB 1401 RNCRLDCSVYKMAEWEECPATCG--THVQOSRNVTCVSAEDGGRTILKDYDC-DVQKRP 1458

QY 838 DPFCSSSGYIQEKCVIPCP-FDCKLSDW--SSNGSCSSSCGIGVIRSRKWLKPKPYNGG 894  
DB 1459 SARNCR-----LEPCPAGEEHIGSWITGDWASKCSASCGGWRARSVSCTSSCDET 1509

QY 895 RCPKXLDLKNQVHEAVPCYSECNQ-----YSWVVEHWSCKINNELSLRCGGGTQS 946  
DB 1510 RK-PKM-----FDKNEBELCPPLTNWQISPWTHC-----SVSCGGGVQR 1549

QY 947 RKIRCVTADGEGGAVDSNLCNQBDEI-----PPTQSC-----SL 981  
DB 1550 RKIWCEDVLSGR-----KQDDIECSEIKPREQRCDEMPPCRSHYHNKTSASMTSL 1600

QY 982 MCPNECVMS-----EW--GLWSKCQSCDPHTMQRRTHLLPSLN-----S 1021  
DB 1601 SSSNNTTSSASASLPLPPVWSNQTSAWSACSAKCGRGT-KRRVVECVNPSLNVTVAS 1659

QY 1022 RTCAEDSOVQPC-----LLNENCFOYNLTWSTCOLSENAPCGQGVTRILLSCVCSDGK 1077  
DB 1660 TEC--DQTKKPVVEVRCRTKHC--PRWKTITWSSCSVT---CGRIGRRREVQCVGRGN 1711

QY 1078 PYSMDQCEHNLKPKQMSIPLVECV-VNCLSGW--TAWTECSQTCGHGRMSRTRFI 1134  
DB 1712 LVSDSECN-----PKTKLSVANCFPVACPAYRWNVTPWSKCKDECARQKQTRRVHC 1764

QY 1135 IMPTQCEGRPCPTTELTOEKT---CPVTPC-YSWVLGNWSACKLEGGDCGEGVQIRSLSC 1189  
DB 1765 ISTSGKRAAPRMCELARAPTIRSICDTSNCPYEWVPGDMQTC---SKSCGEGVQTRVRC 1821

QY 1190 --MVHSGISHAAGVEDALCGEMPFODSLILKQCSV-PCPGD---CHLT-----EWS- 1236  
DB 1822 RKKNFTSTIPTIFLMLED-----EPAPVKEKCELPFKPNESQTCELNCPDSEFKWSF 1873

QY 1237 -EWSFCELTJC-----IDGRSFETV 1254  
DB 1874 GPWGCCKNCGQIRRRRVKCVANDGRVERV 1905

RESULT 7  
US-09-800-729-155  
; Sequence 155, Application US/09800729  
; Patent No. US20020068319A1  
; GENERAL INFORMATION:  
; APPLICANT: Ni et al.  
; TITLE OF INVENTION: 32 Human secreted proteins  
; FILE REFERENCE: P2044P1  
; CURRENT APPLICATION NUMBER: US/09/800,729  
; PRIOR FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: PCT/US00/26013  
; PRIOR FILING DATE: 2000-09-22  
; PRIOR APPLICATION NUMBER: 60/155,709  
; PRIOR FILING DATE: 1999-09-24  
; NUMBER OF SEQ ID NOS: 217  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 155  
; LENGTH: 2165  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-800-729-155

Query Match 8.4%; Score 694; DB 9; Length 2165;  
Best Local Similarity 22.9%; Pred. No. 5e-45;  
Matches 351; Conservative 133; Mismatches 556; Indels 492; Gaps 84;

QY 18 CEHFALQPTQACILPCPR-----DCVSEFLPWSNCKGCKKQHRTRAVIAP-PLF 71  
DB 586 CHH-----GACVRLAPESLTKIDQGWDRWSGECSTRCTGGVQKGLRDCDSPKRN 637

QY 72 GGLQCPNLTES-RACDAPISCPLEEYTFSLKVGPMWKRLPHLKEINPSGRTVLDPNS 130  
DB 638 GGYKCVGQGRYRSCNTQ-ECFMDTQPY-----REVQCS-----EFN- 673

QY 131 DSNERTFKHOSYAHHSKSWAIEIGYQTRQVSCSTRSDGQNALSL-----CLQD 181

674 --NKDIGQVASTNTHWPKYANVAPNERCKLYC-RLSGSAFYLLRDKVVDGTPCDRN 730  
182 SPLTLVQSCIMPDCOE-----TSQSSWSPCSCTCRSGSLLPG----- 219  
731 GDDICVACAGPAGCDHQLHSLRDXKCGVCGDDSSCKVKGTFNEOGTGYNEVMKIP 790  
220 FRSRNRVKNHAIAGGKCPPELLEKEACIVGELLQOCPFRYSWRTSEWKECOVSLLEQ- 278  
791 AGSANIDIRQKYNMKEDDNYLSRA--ANGEFLN-----GHFQVSLARQOI 837  
279 --QD-----PHWHVTGP-----VCGGGIQTREV--YCAQSVRAAALRA 313  
838 AFQDTVLVSGSDAILIERINGTGPFRSDIYVHVLVSGSHPPDISYEVMTAAVPA-- 892  
314 KEVSPVSKAL-----CVGPAP-----LPSQ- 334  
893 --VIRPISALYLWRTDTWTECDRACRQOQSKLMLDMSTHRQSHDRNQNLVKPKQA 950  
335 --LCNIPGCTDGVSSWSAWGLCIHENCHEPQKKGFRTRQRBHVLMESTGPAGCHPLVE 392  
951 TRMCMIDCSTWITEDVS-----SCSAKCGSQKQRQVSCVMMEGDRQTPASEHLCD 1002  
393 --SVPCDEPMCY-----RWLASSEGI-CFPDHGKGLGHLIRLKAQVQNDRGEDVSGSLC- 442  
1003 RNSKPSDIASCVIDCSGRKNYGEWTSCTCTGSGNGKHE--KSYCVDDSNRRYDESLOG 1060  
443 --PVPPPPERKSCIEPCRDVULSEWT--EWSGCSQSNKNSDKQKTRSRITILALAGE-- 497  
1061 REQKEATRECNRIEPC-----PRVYGHWSECSRCDG-----GYQRHAQCLDAADRET 1110  
498 GKGPQPPQALQHELRCLNDHSCMQLHWETSPWGPCSEDTLVTALNATIGWNGEATCGVI 557  
1111 HTRSGPQAT-QEH-CNEHAC--TWAQFVMSDCS-----AKCGDGV 1148  
558 QTRRVFCVSHVQVWTKRCPDSTPRTVPCFLPKKDCIVTAPSEWTPCPRMCQ----- 613  
1149 QYRDANCNTDRSRVLPFRCLKMKEXIIT-KPCF---RESCKYKLGWSQCSVSCDWS 1204  
614 ---AGNAT---VKOSRYEII1QEAANGQCEP----- 639  
1205 SRRVSCVSGNGTEVDMSCIGTASDPASHQTCNLG-TCFWRNTDMSACSVSCGIGHRR 1263  
640 --DTLYERECEDVSLC-----PVYRWKPKQWSPCLLPVPSWQGITSS 682  
1264 TTECIYREQSV-DASFCGDTKMPETSQTCHELLPCTSWKPSHWSFC-----S 1308  
683 EACGKGLQTRAVSC-----ISDDNRSABWMB-CLKQTNHPLLVQECTVPQRE 729  
1309 VTCGSGIQTRSVSCTRGSEGTIVDEYFCDRNTRPRLKTKCTCKDTCDSRVLQK---LQA 1364  
730 DCTFTAWS--KFTPGCTNCEATKSRRLQITGKSRKKEKCD-----SDLYPLVET 777  
1365 DVPPIRMATGPWTACATCG-----NGTQRELLKCRDHRDLDPDEYCNHLDKVEST 1415  
778 ELPCDCEIFISQYGNWSDCIIPEGRBRPHRGLRVQADSKGEGEGLRFAVACSKNGRPV 837  
1416 RNCRLDCSYWKAEMWECAPATCG--THVQQRNVTCSAEDGGRTILKDVDC-DVQXRP 1473  
838 DPFCSGSGYIQEKVIFCP-FCDKLSDW--SSWGSQSSCGIGVIRSRKWLKEKPYNGG 894  
1474 SARNCR-----LEPCPKGEHIGSWITGDKSKCSACGGWRRRSVCTSSSDEET 1524  
895 RCPCKLDLKNQVHEAVPCYBCNQ-----YSWVEHWSCKINNELSLRCLGCGTQS 946  
1525 RK-PRV-----FDKNEBELCPPLTNNSWQISPWTHC-----SVSCGGGVQR 1564  
947 RKIRCNTADGEGGAVDSNLCNDBEI-----PRTQSC-----SL 981  
1565 RKINCEVLSGR-----KODIEGSEIKPRQRCDEMPPCRSHVHNKTSASMTSL 1615  
982 MCPNECVMS-----EW--GLWSKCPQSCDPHTWQRTRHLLPSSLN-----S 1021  
1616 SSSNNTTSSASASSLPILPVVWSQTSAMWSACSAKCGRGT-KRRVVECVNPSLNVIVAS 1674

QY 1022 RTCAEDSOVQPC-----LLNENCFCQFOYNLTENSTCOLSENAPCGQGVTRTELLSCVSCDK 1077  
Db 1675 TEC--DOTXKPBVEVRCRTHKC--PAWKTTTWSGCVT-----CGRIRREVCQTRGRKN 1726  
QY 1078 PYSMDQCEQHNLEKPMQSIPLCEVCV--YNCQLSGH--TAMTECSQTCGHGGRMSRTFI 1134  
Db 1727 LVSDSECN-----PKTKLNSVANCFPVACPAYRWNVTPWSKCKDEACARGQKQTRRVHC 1779  
QY 1135 IMPQTGEGRCPPELTQEXT---CPVTPC-YSWVLGNWSACKLEGDCGEGVQTRSLSC 1189  
Db 1780 ISTSGRAAPRMCELARAPTSIRECDTSNCPYEWVPGDWQTC---SKSCGEGVQTRVRC 1836  
QY 1190 --MVHSGSISHAAGRYEDALCGEMPFQDSILKQLCSV-PCPGD---CHLT-----EWS- 1236  
Db 1837 RRKINFNSTIPIFMLED-----EPAVPKCKELPPKPNESQTCELNCPDSEBFFKNSF 1888  
QY 1237 -EWSCTELTC-----IDORSFETV 1254  
Db 1889 GPMGECCKNGCGQIRRRRVKCVANDGRRVERV 1920

RESULT 8  
US-10-274-639-10  
; Sequence 10, Application US/10274639  
; Publication No. US20030232349A1  
; GENERAL INFORMATION:  
; APPLICANT: INCYTE GENOMICS, INC.  
; APPLICANT: DELEGEANE, Angelo M.; GANDHI, Ameena R.  
; APPLICANT: HAFALIA, April J.A.; LU, Dylung Aina M.  
; APPLICANT: PATTERSON, Chandra; TRIBOULEY, Catherine M.  
; APPLICANT: DAS, Debopriya; KALLICK, Deborah A.  
; APPLICANT: NGUYEN, Daniel B.; LEE, Ernestine A.  
; APPLICANT: KHAN, Farrah A.; YUE, Henry  
; APPLICANT: AU-YOUNG, Jennifer K.; GRIFFIN, Jennifer A.  
; APPLICANT: POLICKY, Janice L.; RAMKUMAR, Jayalaxmi  
; APPLICANT: YANG, Junming; THANGAVELU, Kavitha  
; APPLICANT: DING, Li; KEARNEY, Liam  
; APPLICANT: BAUGHN, Mariah R.; BOROWSKY, Mark L.  
; APPLICANT: SANJANWALA, Madhusudan M.; YAO, Monique G.  
; APPLICANT: BURFORD, Neil; WALIA, Narinder K.  
; APPLICANT: LAL, Preeti G.; LEE, Sally  
; APPLICANT: TODD, Stephen; LO, Terence P.  
; APPLICANT: TANG, Y. Tom; ELLIOTT, Vicki S.  
; APPLICANT: AZIMZAI, Valda; LU, Yan  
; TITLE OF INVENTION: PROTEASES  
; FILE REFERENCE: PI-0167 USA  
; CURRENT APPLICATION NUMBER: US/10/274,639  
; CURRENT FILING DATE: 2002-10-18  
; PRIOR APPLICATION NUMBER: PCT/US01/22397  
; PRIOR FILING DATE: 2001-07-17  
; PRIOR APPLICATION NUMBER: US 60/220,063  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: US 60/221,680  
; PRIOR FILING DATE: 2000-07-28  
; PRIOR APPLICATION NUMBER: US 60/223,544  
; PRIOR FILING DATE: 2000-08-04  
; PRIOR APPLICATION NUMBER: US 60/224,717  
; PRIOR FILING DATE: 2000-08-11  
; PRIOR APPLICATION NUMBER: US 60/225,986  
; PRIOR FILING DATE: 2000-08-16  
; PRIOR APPLICATION NUMBER: US 60/227,568  
; PRIOR FILING DATE: 2000-08-23  
; NUMBER OF SEQ ID NOS: 42  
; SOFTWARE: PERL Program  
; SEQ ID NO 10  
; LENGTH: 1916  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. US20030232349A1 2994162CDI  
US-10-274-639-10

Query Match		8.2%; Score 674; DB 15; Length 1916;
Best Local Similarity		21.4%; Pred. No. 1.9e-43;
Matches		290; Conservative 123; Mismatches 419; Indels 526; Gaps 70;
QY	201 WSS---WSPCKTRSGSLLPGRFSRNRNVKHMAGGKCEPPELLEKACIVEGELL---	254
Db	862 WNSHGPWQACSKPCQ-----GERKR-----KLVTRESDLQTVS	895
QY	255 -QCCPRYSWRTSEWKEQVSLLEEQDPHWHVT-----GPVCGGIGTREVYCAQSVPA	308
Db	896 DQCDRLPQPHITEPCGTDG-----DLRHWASRSECSAQCGLYRTLDIYCAK-----	945
QY	309 AALRAKEVSRVEKALCVGPAPLPSCNIPCTDCIVSSW--SANGLCIHENCHEPQGX	366
Db	946 -YSRLDGKTEKVDGDFC--SSHPEKPSN--REKSGECNTGGWRYSAWTEC-----	990
QY	367 KGFRTQRQHVLMESTGPAGHCPHLVESVPCEDPMCYRWLASGICPDHKGCGLGHRLK	426
Db	991 -----SKSCDGGTQRRR	1002
QY	427 AVCQNDRGEDVSGSLCPVPPPPERKSC-EIPCRMDCVLSEWT--EWSSCSQSCSNKNSDG	483
Db	1003 AICVNTRENDVLDDSKTHQEKVTTQRCSEFFC-----PQWKGDMSECLVTC-----	1050
QY	484 KQTESRTILALAGE---GGKPCPSQALQEHRLNDHSCMQLHETSPPWGPCSDETLVTA	540
Db	1051 KGHKRWQCOFGEDRLNDRMCDPETHPTSMQTCQPECAS--WQAGPWGQCS-----	1101
QY	541 LNAITWNGEATCGVIGTQTRVFC-VKSHVQVQVTKRCPDSTRTPETVRPCLFCKKDC--	597
Db	1102 -----VTCQGGYQLRAVAKCIIGTVMSVDDNDCAATRTPTDQCELP--SCHP	1148
QY	598 -----IVTAFSEWTPCPRCQAGNATVKSRYRIIOEAANGQGCPTD	641
Db	1149 PPAAPETRSTYSAPRTQWRFGSTPCSATCGKG-----TRMYVSCRDENG-----	1196
QY	642 LYBEREC-----EDVSLCPYRWKPKQKWSCILVPESVWQGTGSGEACGKGLQTR	692
Db	1197 VADSSACATLPRVAKKECSVTPCGQKALDWSC-----SVTCQGGRAIR	1242
QY	693 AVSCISDNRSAEMWELCKQTNMGMLLVQSCVPCREDCTFTAWSKPTPSTNCEATKSR	752
Db	1243 QVMCVNTSDHVIDRSEC--DQDYIPKTDQCSM-----SPC-----	1276
QY	753 RRLQTKSRKKEKQDSDL--YPLVETELCPDEFISQPY-----GNWSDCILPEG	801
Db	1277 -----PQRTPDGLAQHPQONEDYPRSPASRTHVLGNQWRTPGWAC-----	1321
QY	802 RREPHRLRVOADSEKCEGLRFRVAVACDKNG-----RPVDPSPCSSSGYIQEX	851
Db	1322 -----SSTCAGGSQRRVVVQDENGVTANDCVRIKPDQRACESG-----	1362
QY	852 CVIPCFDCKLSDWS--SWGSCSSCGIGVIRISKWLKEXPYNGRPPCKPLDKNQVHEA	909
Db	1363 ---PCP-----QWYGNWGETKLCGGGRTR--LVVQQRNGERFP--DUSCSILDK	1408
QY	910 VPCYSECNQY-----SWVYEHWSCKINNELSLRGGGTQSRKIRCVNTADGEGGAVD	963
Db	1409 PPDEQCNTHACPHDAKSTGPMWSSC-----SVSCGRGHQNRVYCMK-----	1458
QY	964 SNLCNQBEPETQSC-SLMCPNECVWSEW--GLWSKCPQSCDPHTNQRTHLLRPSLN	1020
Db	1459 SDYCKHLAKPHGRKCRGGRCP-----KWKAGAWSQCSVSCRGVQQRH-----	1508
QY	1021 SRTCAEDSQVQCLLNEN-----CFQFOYNLTMTSTQOLSENAPCGQGVTRLLSCV	1072
Db	1509 THKTARETECNPTRPESERDCQGRCPFLYTWRAEWQEC-----TKTCGGSRYRKVCV	1564
QY	1073 -----CSDGK--PVSMQDCEQHNLEKFORMSTIPCLVECVNQCLSGWTA--WTEC	1118
Db	1565 DDNKNVHGARDVSKRPVDRSCLQ-----PC-EYV-----WITGEWSEC	1605
RESULT 9		
US-09-938-330-25		
; Sequence 25, Application US/09938330		
; Patent No. US20020115838A1		
; GENERAL INFORMATION:		
; APPLICANT: Walke, D. Wade		
; APPLICANT: Hilbun, Erin		
; APPLICANT: Scoville, John		
; APPLICANT: Friedle, Carl Johan		
; APPLICANT: Hu, Yi		
; APPLICANT: Turner, C. Alexander Jr.		
; TITLE OF INVENTION: No. US20020115838A1el Human Proteases and Polynucleotides Encodin		
; FILE REFERENCE: LEX-0237-USA		
; CURRENT APPLICATION NUMBER: US/09/938,330		
; CURRENT FILING DATE: 2001-08-22		
; PRIOR APPLICATION NUMBER: US 60/227,104		
; PRIOR FILING DATE: 2000-08-22		
; PRIOR APPLICATION NUMBER: US 60/233,796		
; PRIOR FILING DATE: 2000-09-19		
; NUMBER OF SEQ ID NOS: 26		
; SOFTWARE: FastSeq for Windows Version 4.0		
; SEQ ID NO 25		
; LENGTH: 1907		
; TYPE: PRT		
; ORGANISM: homo sapiens		
US-09-938-330-25		
Query Match		
8.2%; Score 673; DB 9; Length 1907;		
Best Local Similarity		
21.4%; Pred. No. 1.9e-43;		
Matches		
290; Conservative 123; Mismatches 419; Indels 526; Gaps 70;		
QY	201 WSS---WSPCKTRSGSLLPGRFSRNRNVKHMAGGKCEPPELLEKACIVEGELL---	254
Db	853 WNSHGPWQACSKPCQ-----GERKR-----KLVTRESDLQTVS	886
QY	255 -QCCPRYSWRTSEWKEQVSLLEEQDPHWHVT-----GPVCGGIGTREVYCAQSVPA	308
Db	887 DQCDRLPQPHITEPCGTDG-----DLRHWASRSECSAQCGLYRTLDIYCAK-----	936
QY	309 AALRAKEVSRVEKALCVGPAPLPSCNIPCTDCIVSSW--SANGLCIHENCHEPQGX	366
Db	937 -YSRLDGKTEKVDGDFC--SSHPEKPSN--REKSGECNTGGWRYSAWTEC-----	981
QY	367 KGFRTQRQHVLMESTGPAGHCPHLVESVPCEDPMCYRWLASGICFPDHKGCGLGHRLK	426
Db	982 -----SKSCDGGTQRRR	993
QY	427 AVCQNDRGEDVSGSLCPVPPPPERKSC-EIPCRMDCVLSEWT--EWSSCSQSCSNKNSDG	483



b 994 AICVNRNDVLDSDCKTHQKVTIQRCEFFPC-----POWKSQDMSCLVTC-----G 1041  
y 484 KOTRSRTIILALAGE---GGKPCPPSQALQEHRLCNHDSQMLHWTSPWGPCSBDTLVTA 540  
b 1042 KGHKRWQVQCFGEURLNDRMCDPTKTSMTCCQPECAS---WQAGPWGQS----- 1092  
y 541 LNATIGWNGEATCGVIGIOTRRVFC-VKSHVGQVMTKRCPSDSTRPTVRPCLPKCKDC-- 597  
b 1093 -----VTCGGGYQLRAVKCIIGTYMSVVDNDNCNAATRPDTQDCCLP---SCHP 1139  
y 598 -----IVTAFSEWTPCPRMCOAGNATVKQSRVRIIIQEAANGQCEPDT 641  
b 1140 PPAAPETRRSTYSAPRTQWRFGSWTPCSATCKG-----TMRVSVCRDENG-----S 1187  
y 642 LYBEREC-----EDVSLCPVYRWKFPQKWSPCILVPESVMOGITSSACGKGLQTR 692  
b 1188 VADESACATLPRPAKECSVTPCGWKALDWSSC-----SVTCGGGRATR 1233  
y 693 AVSCISDDNRSAMMECLKQTNMGMLLVQECTVPCREDCTFTAMSKFTPCSTNCEATKSR 752  
b 1234 QMVCNYSHVHIDRSEC---DQYIIPETDQDCSM-----SPC----- 1267  
y 753 RRLTGKSRKEKQCDSDL--YPLVETELPCDEFISQPY-----GNWSCILPEG 801  
b 1268 -----PORTDPSGLAHPFQNEIDYRPSASPSRTHVLGGNQWRTGPWAC----- 1312  
y 802 RREPHRGLRVQADSKEGEGLEFRVAVASDKNG-----RPVDSFCSGGYQEK 851  
b 1313 -----SSTCAGGSQRRVVVQDENGTYTANDCVERIKPDEQACESG----- 1353  
y 852 CVIPCPFDCKLSDWS--SWGSCSSCGIGVRLSRKWLKEKPVNGRCPKPLDLKNQVHEA 909  
b 1354 ---PCP-----QWAGNWGECKLGGGIRTR---LVVQSRNGERP---DLSCILDK 1399  
y 910 VPCYSECNOY-----SMVVEHWSCKINNELRLRCGGGTQSRKIRCVNTADGEGAVD 963  
b 1400 PPDREQCNTACPHDAANSTGPWSSC-----SVSCGRGHQKRVNYCMAK---DGSHE 1449  
y 964 SNLQNDQEIPEPQSC-SLMCPNECVMSBW--GLWSKCPQSCDPTHQWRTHLLRLSLN 1020  
b 1450 SDYCKHLAKPHGRKCRGRCP-----KWAGAWSQSVSCGRGVQORH---VGCOIG 1499  
y 1021 SRTCAEDSQVOFCLLEN--CFQFOYNLTWSTCOLSENAPCGQGVRTLLSCV 1072  
b 1500 THKIARETECNPYTRPESERDCQGPRLTYWRAEWQEC---TKTCGEGSRVYKVVCV 1555  
y 1073 -----CSDGK-PVSDMOCEQHNLEKPRMSIPCLVECVNQLSGWTA--WTEC 1118  
b 1556 DNKNVEHGARDVSKRPVDRSCSLQ-----PC--EYV-----WITGEWSEC 1596  
y 1119 SOTCGHGRMSRTRFIIMPTQGEGRPCPELTQEKTI--CP-----VTPCY-----S 1162  
b 1597 SVTCGKYQKRLVSCSEIYTKEN---YESYQYTTINCPGTQPPSHVPCVLRCPVSAT 1652  
y 1163 WVLGNWSACKLEGGCGGVQIRLSUSVMVHSGSISHAA----- 1200  
b 1653 WRVGNWGSV---SCGVGMQSRVQCLTNEDQPSHLCHTLKPEERKTCRNVCNCLPQ 1709  
y 1201 -----GRVEDALCE--MPFQDSILKQLCSVPFCDCHLTWSENSCELTCDGR 1249  
b 1710 NCKEVRKLKGAED---GEYFLMIRGKLKLFICA-----GMHSDHPKRYVT--LVHGDSE 1759  
y 1250 SF-ETVGRQSRRTFIIOSEFNQDSC-----POQVLET----- 1281  
b 1760 NPEYVGHRLHNTEPCYNGSRDRDCQCKDYTAAGFSSFKIRIDLTSMIITDQLQA 1819  
y 1282 ----RP---CTGGKCYHT-----WKASLWN-----NERTVWCQS-----D 1312  
b 1820 RTSEGHVPFFATAGCYSAKCPQGRFSINLYGTGLSLTESARVISQGNVAVSDIKKSPD 1879  
y 1313 GVNVTGGGSPQARPAAIRQICPACRKPESYCTQGGVCG 1350

Db 1880 GTRVVGKC-----GGYCG 1892  
RESULT 10  
US-09-918-171A-13  
; Sequence 13, Application US/09918171A  
; Patent No. US20020110894A1  
; GENERAL INFORMATION:  
; APPLICANT: Apte, Suneel  
; APPLICANT: Hurskainen, Tiina L.  
; APPLICANT: Hirohata, Satoshi  
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases  
; FILE REFERENCE: 26473/04193  
; CURRENT APPLICATION NUMBER: US/09/918,171A  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: 09/369,364  
; PRIOR FILING DATE: 1999-08-06  
; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 1882  
; TYPE: PRT  
; ORGANISM: Homo sapiens ADAMTS-9  
; FEATURE:  
; NAME/KEY: MOD RES  
; LOCATION: (468)  
; OTHER INFORMATION: Xaa = Cys  
; NAME/KEY: MOD RES  
; LOCATION: (521)  
; OTHER INFORMATION: Xaa = Tyr  
US-09-918-171A-13  
Query Match 8.0%; Score 657; DB 9; Length 1882;  
Best Local Similarity 21.2%; Pred. No. 3.3e-42;  
Matches 288; Conservative 123; Mismatches 421; Indels 526; Gaps 70;  
QY 201 WSS---WSPCKTCRSGSLPGFRSRNRVKNEMAGGKCEPHELLEKEACIVEGELL--- 254  
Db 828 WNSHGWQACSPQ-----GERKE-----KLVTRESQJTVS 861  
QY 255 -QCPRYSWRTSEWKECQVSLLEOODPHMVT-----GPGCGGIQTRVYCAQSVPA 308  
Db 862 DQRCRLPQFGHITPCGTGC-----DLRHWASRSECSAQGLGYRTLDIYCAK----- 911  
QY 309 AALRAKEVSRPEKALCVGPAPLPSOLCNIPGSTCIVSSW---SANGLCIHENCHPEQK 366  
Db 912 -YSRLDGKTEKVDGFC--SSHKPSN--REKCSGECNCGWRYSAWTEC----- 956  
QY 367 KGFRTQRHVLMESTGPHGCHPLVESVPCEDPMCVRWLASEGICFPDHGKGLGHRILK 426  
Db 957 -----SKSCDGGTQRRR 968  
QY 427 AVCONDRGEDVSGSLCPVPPPERKSC-EIPCRMDCVLSEWT---FWSSCSQSCSNKSDG 483  
Db 969 AICVNRNDVLDSDCKTHQKVTIQRCEFFPC-----POWKSQDMSCLVTC-----G 1016  
QY 484 KOTRSRTIILALAGE---GGKPCPPSQALQEHRLCNHDSQMLHWTSPWGPCSBDTLVTA 540  
Db 1017 KGHKRWQVQCFGEURLNDRMCDPTKTSMTCCQPECAS---WQAGPWGQS----- 1067  
QY 541 LNATIGWNGEATCGVIGIOTRRVFC-VKSHVGQVMTKRCPSDSTRPTVRPCLPKCKDC-- 597  
Db 1068 -----VTCGGGYQLRAVKCIIGTYMSVVDNDNCNAATRPDTQDCCLP---SCHP 1114  
QY 598 -----IVTAFSEWTPCPRMCOAGNATVKQSRVRIIIQEAANGQCEPDT 641  
Db 1115 PPAAPETRRSTYSAPRTQWRFGSWTPCSATCKG-----TMRVSVCRDENG-----S 1162  
QY 642 LYBEREC-----EDVSLCPVYRWKFPQKWSPCILVPESVMOGITSSACGKGLQTR 692  
Db 1163 VADESACATLPRPAKECSVTPCGWKALDWSSC-----SVTCGGGRATR 1208  
QY 693 AVSCISDDNRSAMMECLKQTNMGMLLVQECTVPCREDCTFTAMSKFTPCSTNCEATKSR 752

```

; TYPE: PRT
; ORGANISM: H. sapiens
Us-10-213-509-5

Query Match      8 0%; Score 655.5; DB 14; Length 4123;
Best Local Similarity 21.9%; Pred. No. 1.2e-41;
Matches 390; Conservative 149; Mismatches 567; Indels 677; Gaps 101;

QY 28 EQACLLPCPRDCVVSFELPWSNCSGCKKQLHRTRAVIAPPFLGGLOCP-NLTESRACD 86
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2507 EDGCV-----DCVLAFWSVMSCSRSGCLGLTFQRELLRPPLPGG-SCPRDRFRSQCF 2560
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
87 APISCPGLBEEYTFSLKVPWSKRLPHLKEINPSGRTVLDFNSDSNERVTFKHQSYKAH 146
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2561 VQ-ACV-----AGAWAWNAWGPCSVCGG-----2565
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
147 HHSKSWAIBIGYQTVQSVCTSRDQGN--AMLSLCLQDSFPLTVQSGIMPDKDCTSOWS-- 202
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2586 -----GHQSRQSRSCVDPKPKNGGAPCPGASQERAPCGLQPCSGGKTGKVLGWGHG 2635
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
203 -----SW-----SPCSKTCRS-----GSLLPGRFSRSRNVK 228
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2636 GSTVGTGRGLGAPAPRLTWCPSTRLRAGPCVCRSPGEGAGAPMTLLPG--SQGQKL 2693
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
229 HMAIGG-----KCPPELLLEKE-----245
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2694 QWALCGSLPSLLCPLGLSALFHLPGCRCPPLLHLDTRCLPLSECPCLGVEELKWPVGS 2753
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
246 -----ACTIVE-HELLQ--CP-RYSWRT-SEWKECQVSLLEQODPHHWTVGPVCGG 292
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2754 FLIENGSCQVCERKELLCPGGCFLPCGMSAWSSWAPCDAS-----CGS 2797
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
293 GIDTREVYCAQSVPAALRAK-EVSRP-----VERALCVGPAPLPSOLCNIP 339
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2798 GVRR--FRSPNPPAAWGAPCEGRDRELQGHCTVCGTGIAGSLGAVPPSSQFTCLR 2855
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
340 C-----STDCIV-----SSWSAWGLCIHENCHEPQKKGFRTRQHVLMESTGPAGHC 387
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2856 THGMGPTDHTSTGIEVFGWTPMTWSSC-SQSLAPGGGPGWRSR--LCPSGDSS-C 2911
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
388 P-HLVSVPCEDPW-CYRWLASGICFPD-----HGK-----417
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2912 PGDATGEPCSPPIECTGFCAPGCTCPGLFLHNASCLPRSQCPCQLHGLQIYAGAWARL 2971
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
418 -----CGLGHRILKAVCONDRGEDVSGSLCVPVPPPERKSCIEIFCRMDCVLSEWTEWS 470
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2972 DSCNNCTCVSG---KMACTSER-----CFVA-----CGWSPWLWS 3004
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
471 SCQSQCSNKNNSDGKQTRSRTILA-LAGEGKPCPPPSQALQEHRLCHNDHSC-----M 520
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3005 LCSCSC---NVGIRRFRRAGTAGTAPPAFGAEC--QGPTWEAFSCSLRCPGVPVGMCPR 3058
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
521 QLHWETSPWGP--CSEDTLVTALNAT-----545
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3059 DKQWLDAQGPASCABELSAPRGNTQICHPCCHPSGMILLVSRGHPPGLGASVQPPVAL 3118
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
546 -----GW-----NGEATCGVGIOTRRVFC-----VKSHV 569
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3119 PGAIGTGSVEGAGWGPWSPWHSRCSRGGLRSRTRACDPFPQGLGDYCEGPRAGQEV 3178
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
570 GOVWTKRCP--DSTRPETHR--PCFLPCKKDCIVTAFSEWTPCPRMCQAGNATVKQRYR 625
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3179 QCALP--CPVNTCTATEGAEYSFGPPCPRSCDDLVEHCVR--QCPGCYCPGGQVLSNGA 3235
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
626 IITQEAANGQECFDPITYER-----ECEDVSLCPVYRWKQP 662
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3236 ICVQP---GHCSCLDLITGQRHHPGARLARPDGCHNCTCLEGRINCTDLR-CPDC-GGGQ 3290
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
663 KWSPC-----ILVPESVMO-GITGSEACG-----XGLQTRAVSCISDDNRSA 704
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3291 SLHPGQCPFRSCQDLSPGSCVQPGSVGQOPTCGPLQLGSLQDGLCVPPAHM-----3342
QY      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
705 EMMECLKQTINGM-PLLVQECTV-----PCRE--DC---TFTAWSKTIPTCSNLC---EATK 750

```

Qy	546	----	GW	-----	NGEATCGVGIQTRRVFC	-----	VKSHV	569									
Db	3119	PGAIGTSVEGAGGWP	GWP	SWSCRS	CGGLRSRTRACD	QPPQPQGLDGYCEGPRAQCEV	3178										
Qy	570	GOVMTKRCP	--	DSTR	RETVR	--	PCF	PCKDCIVTAFSEWTPCPRVQAGNATVKQSYR	625								
Db	3179	QALP	--	CPVT	NCIALTEGARYSP	CGPPCPRSCD	DLVHCVMR	--	QPGCYCPGPGQLSNGA	3235							
Qy	626	IIIQEAANGQEC	PD	TLYBER	-----	ECEDVSLCPYRWKQP	662										
Db	3236	ICVQ	---	GHC	CLDLLTGORHH	PGARLARPD	GCNCTCLEGR	IN	CTDLP	--	CPDC	GGGQ	3290				
Qy	663	KWSPC	-----	ILVP	SVWQ	--	GITG	SEAG	-----	--	KGLQTRV	ASCISDDN	RESA	704			
Db	3291	SLHPCQCP	PR	SCQD	LS	PGSV	CGP	SGV	CGP	TCG	PLG	LSQDGL	CVPPAH	-----	3342		
Qy	705	EMMECLQ	TNGM	--	PLIV	QECTV	----	PCRE	--	DC	----	TFTAWSK	FTPC	STNC	----	EATK	750

b 3343 ----RCYQGAAPSFVSTCVAGILQCOEVPDPPGVMWSWGFWDCSVSCGGEQLR 3399  
y 751 SRR-----RLTGSRKKEKQDSDLYPLVET-ELPCD-EFISQPVGNWSD 795  
b 3400 SRRCARPPCGPARQRTCTQVCREAGCPAGRLYREQPGEGCFSCAHVTQVQGCSE 3459  
y 796 ----CILPEGRERHRLVQ-----ADSKGEGGLR 823  
b 3460 GCEGCHCEGTFQ-HRLACVQECPCVLTAMLLQELGATIGDPGOLPGDELDSGQTLR 3518  
y 824 FRAVACSKNGR---PVPSPFCSSGY-----IQKCVIPCP-- 858  
b 3519 TSCGNCSAHGLKSLDDCFADGFGPWPSPGSCRSGLGTRTSRQCVLTMTPLS 3578  
y 859 ----DKLSDWSWGSCSSCGIGVIRSKWLKEKPYNGRPPCLDLKNQ 905  
b 3579 ELPVCPGPGCGAGNCWSWAPWEPSCRSVGVQOORLRAVR-PPGPGHWCNILLTAYQ 3637  
y 906 VH-----EAVPCYSEGNYSWVVEHWSCKINNEL-RSLRCGGTOSR 947  
b 3638 ERFNCLRACPAAGCPAGNEVVTANRCPRCSDLOEGIVQDDQVQCGKCPKGSLEQ 3697  
y 948 KTRCV-----NTADGEG-----GAVDSNLQDEIPPTQSCSLM-C--PNECYMSEWGL 994  
b 3698 DGGCVPIGHCDCTDAQHSHWAPGSHQDACNNCSQAGLSCTAQPCPPPTHCAWHSWA 3757  
y 995 WSKCPQSCDPTMQRTTRHLLPNSLR--TCAEDSOVQPCLL-----NE 1037  
b 3758 WPCSHSCGPRCQQRFR--CGFGLASGSGC-----PCLMAKADPTCNSTFLHLDTQ 3808  
y 1038 NCF-----QFOYNL--TWSSTQLSNAPCGG--VTRLLSCVCSG--KPVSMDOCEQ 1086  
b 3809 GYSGPCPDSCQWSLWGPSPQV---PCSGFRLRWREAEALCGGCRFPQADR--- 3861  
y 1087 HNLXPQMSIP--CLVRCV---NCLSG-WTAWTECSTQCHGHWSTRIIIMPTQ 1139  
b 3862 ----KLORALPSTCNESLVCPHOECPVLGFWPWSWSSCAPCG-GGTMRHRTC---EG 3913  
y 1140 GBG-RPCPELT-QEKTCPVTPCYSMWLN-WSACKLEGGCGEGVQIRLSLM---VHS 1193  
b 3914 GGVAPCAQADTEQRECNLQPCPCPGQVLSAC-----ATSCPLCWHLQ 3961  
y 1194 GSISHAAGVEDALGEMFPFQSIILKQISVPCPDGCHLTWSEWSTCELTCIDGRSFET 1253  
b 3962 G-----AICVQSPCPG-----CG--CPGGHSLPWGLTILE----- 3992  
y 1254 VGRQSRRTFIIQSFENODSCFQOVLETRPCTGKCYHYTWKASINNNERTVNCQSDG 1313  
b 3993 -----EQAQLPPGTVLTRNCTRCVCHGGAFCSL-----VDCQE-- 4028  
y 1314 VNVGCGSPQAPAAIROCIPAC-----RKPFSTCTGGVCGC 1351  
b 4029 IVPGETWQVAPGELGLCEQTCLMNAKTKQSNCSARASGC 4071

## RESULT 12

US-10-085-198-2

Sequence 2, Application US/10085198

Publication No. US2004000907A1

GENERAL INFORMATION:

APPLICANT: Alsobrook et al.

TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same

FILE REFERENCE: 21402-279

CURRENT APPLICATION NUMBER: US/10/085,198

CURRENT FILING DATE: 2002-02-25

PRIOR APPLICATION NUMBER: 60/271,646

PRIOR FILING DATE: 2001-02-26

PRIOR APPLICATION NUMBER: 60/276,401

PRIOR FILING DATE: 2001-03-16

PRIOR APPLICATION NUMBER: 60/311,981

PRIOR FILING DATE: 2001-08-13

PRIOR APPLICATION NUMBER: 60/312,858

;; PRIOR FILING DATE: 2001-08-16  
;; PRIOR APPLICATION NUMBER: 60/271,840  
;; PRIOR FILING DATE: 2001-02-27  
;; PRIOR APPLICATION NUMBER: 60/277,324  
;; PRIOR FILING DATE: 2001-03-20  
;; PRIOR APPLICATION NUMBER: 60/286,096  
;; PRIOR FILING DATE: 2001-04-21  
;; PRIOR APPLICATION NUMBER: 60/299,695  
;; PRIOR FILING DATE: 2001-06-20  
;; PRIOR APPLICATION NUMBER: 60/315,614  
;; PRIOR FILING DATE: 2001-08-29  
;; PRIOR APPLICATION NUMBER: 60/272,405  
;; PRIOR FILING DATE: 2001-02-28  
;; Remaining Prior Application data removed - See File Wrapper or PALM.  
;; NUMBER OF SEQ ID NOS: 653  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 2  
;; LENGTH: 4219  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; US-10-085-198-2

Query Match 8.0%; Score 655.5; DB 15; Length 4219;  
Best Local Similarity 21.9%; Pred.No.1.3e-41;  
Matches 390; Conservative 149; Mismatches 567; Indels 677; Gaps 101;  
QY 28 EQACLIPCRDCVVSFELPWSNCSKCGKGLQHRTPAVIAPPLFGGLQCP-NLTESRACD 86  
Db 2603 EDGCV-----DCVLAPWSVWSSCSGGLGTFQQLLRPLPGG-SCPRDRFRSQC 2656  
QY 87 APISCPGEBEYTFSLKVPWSKCRPLHLKEINPSRTVLDNFNSDSNERVTFKHQSYKAH 146  
Db 2657 VQ-ACPV-----AGAWAWEAWGPCSVSCG- 2681  
QY 147 HHSKSWAIGVQTRVQSVCTRSQGN--AMLSLQDSPLTVQSCIMPDKCETSQWS-- 202  
Db 2682 -----CHQSRQSCVDPPPKNGGAPCPGASQERAPCGLOPCSGGTGKVLGWGHG 2731  
QY 203 -----SW-----SPCSKTCRS-----GSLLPGRFSRSRNVK 228  
Db 2732 GSTVGTGRGLPAPRLTWCPSTRLPAGFCVCEKSVPEGAGAMPILLPG--SQGQKL 2789  
QY 229 HMAIGG-----KCCPELLEKE----- 245  
Db 2790 QWALCGSLPSLLCPLGLSALFHLPGCRPPGGLLLHTRCLPLSECPCLVGBELKWPVGS 2849  
QY 246 -----ACTIVE-CELLQO---CP-RYSWRT-SEWKEQCVSLLLEQDDPHWHVTGPVCGG 292  
Db 2850 FLGNCSCQCVCKGELLQCPGGCPLPCGWSAWSWAPCDRS-----CGS 2893  
QY 293 GIQTRVYCAQSVPAALALRAK-EVSRP-----VEKALCVGPAPLPQLCNTP 339  
Db 2894 GVRAR--FRSPNPAAWGGAPCEGDRQELQCHTVCGTGTAGSLGAGVPPSSQFCTLR 2951  
QY 340 C-----STDGIV-----SSWSAWGLCIHENCHPQKKGKRTQRHVLMESTGPAGHC 387  
Db 2952 THGMGPTDHTWGIEVFGWTPMTWSSC--SQSLAPGPGGFGWRSR--LCPSPDGSS--C 3007  
QY 388 P-HLVESVPCEDPM-CYRWLASGECIFPD-----HGK----- 417  
Db 3008 PGDATQEEPCSPPIECTGFCAGCTCPGLFLHNASCLPRSCQPCQLHGQLYASGAMARL 3067  
QY 418 -----CGLGHRILKAVQNDRGEDVSGSLCFVPPPPPERKSCBPCRMDCVLVSWTEWS 470  
Db 3068 DSCNNCTCVSG---KMACTSER-----CFVA-----CGWSPWTLWS 3100  
QY 471 SCQSQSNKNSDKOTRSRTILA-LAGEGKGCPCPSQALQEHRLCNHDS- 520  
Db 3101 LCSGSC-----NVGIRRRFRAGTAPPAFGAGSC--CGPTNEAEFCSLRCPGPFVGMCMR 3154  
QY 521 QLHWETSPNGP--CSEDTLVTALNATI----- 545  
Db 3155 DKQWLCAQGPASCAELSAPRGNTQTHPGCHCPSGMLLLVSPRHPGFLGASVQPPVAL 3214

546 ---GW-----NGEATCGVGIOTRREVC-----VKSHV 569  
3215 PGAIGTGVAGGWPWGSWSRSCGGLRSTRACDQPPQGLGBYCEGPRAQGEV 3274  
570 GQVWTKRCP--DSTRPETYR--PCFLPCKKDCIVTAFSEWTPCPKCOAGNATVQSYR 625  
3275 CQALP--CFVNTCTAIGABYS CGPPCRSCDDL VHCVMR--CQPCYCPGPGVLSNGA 3331  
626 I11QEAANGQECPTLYER-----ECEDVSLCPVVRWKPQ 662  
3332 ICQVP---GHCSCLDLTGQRHHPGARLARPDGCHNCTCLEGRNCTDLP-CPDC-GGQ 3386  
663 KNSPC-----ILVPSVMQ-GITGSEAG-----KGLQTRAVSLSDNRSR 704  
3387 SLHPGQPCPRSCODLSPGVCQPGVGOCTCGPLGOLSDGLCVPPAHC----- 3438  
705 EMMECLKQINGM-PLVQECTV-----PCRE--DC-----TFTAWSKTPGCTINC-----EATK 750  
3439 --RCQVQPCGAMAPSFVPTCTVAGILQCOEVPDPPGVWSSWGPWEDCSVSCGGGEQLR 3495  
751 SRR-----RQLTKSRKKEKQDSLYPLVET-ELCPD-C-BFISQPVGNWSD 795  
3496 SRRCARPPCGPARQSRCTOVCREAGCPAGRLYRECQPGEGCPFSCAHVTQOVGCRSE 3555  
796 ---CILPGRREPRLRVQ-----ADSKGEGELR 823  
3556 GBEGCHCEPTEQ-HLACVQBCPCVLTAWLLOELGATIGPGQPLGPDLELDSOTLR 3614  
824 FRVAGSDKNGR--PVDPSFCSSGY-----IOEKCVPICPP-- 858  
3615 TSCGNSCAHGKLSLDDCFEADGGFGPWSWPGPCSRSCGGLGTRRSRQCVTWPTLS 3674  
859 ---DCKLSDWSGSCSSCGIGVIRSRKWLKEKPYNGRGRPCPKDLKNO 905  
3675 ELVPCPGCGGAGNCWTSWAPWEPSCRGVQQRRLRAYR--PPGPGHWPNTLTAQ 3733  
906 V-----EAVFCYSECNQYSVVVEHSSCKINNEL-RSLRCGGGTQSR 947  
3734 ERRFCNLRACEAGCPAGMEVVT-CANRCPRRCDLQEGIVCQDDQVCQKGRCPKGSLEQ 3793  
948 KIRCV-----NTADGG-----CAVDSNLQNOEIPETQSCSLM-C--PNECVMSWGL 994  
3794 DGGCVPIGHCDCTDAQSHWAPGSHQDANNCSCQAGLSCTAQCPPTTHCAWHSR 3853  
995 WSKPOSQCDHTWQRRTHLLRSLNSR--TCAEDSQVQPCLL-----NE 1037  
3854 WSPCSHSGPRGQSRFR--CGFLASRSGC-----PCLMAKADPTCSNFTLHLDQ 3904  
1038 NCF-----QFQVNL-TEWSTCOLSENAFCQG--VTRLLSCVCSG--KPVSMQCEQ 1086  
3905 GCYSGPCPDSQWLSMPWSPQV-----PCSGGFLRWREALCGGCRFPWAQDR--- 3957  
1087 HNLEKFORMIP--CLVECVV-----NCQLSG-WTAWTECSQTCGHGRMSRRTFTIMPTQ 1139  
3958 ---KLORALPCTCVNESLVCBQECVPLGPHSAWSSCSAPCG-GGTWERHTC---EG 4009  
1140 GEG-RPCPTLT--QKTCPTVTPCYSWVLGN-WBACKLEGGDCGQVQIRLSLM---VHS 1193  
4010 GPGVAPCAQDTQRCQFCNLQPCFPCPPGOVLSAC-----ATSCPLCWHLQ 4057  
1194 GSI SHAAGVEDALCEMPQDSILKQLCSVPCPGDCHLTWSENSTCELTICIDGRSPET 1253  
4058 G-----AICVQEPQCG-----CG--CPGQHSPLWGLITLLE----- 4088  
1254 VGRQSRRTFIIOSFNQDSQCPQVLETRPCTGGKYHYTWKASLWNNNRTVWCQRSDG 1313  
4089 -----EQABLPGTILTRNCTRCVCHGGAFCSL-----VDQGE-- 4124  
1314 VNYTGGSCPOARPAARQCIAPAC-----BKPSYCTQGGVCCG 1351  
4125 IVPFGTWOQVAFGELGELCBQTCLEMNATKTONCSARSAGC 4167

RESULT 13  
US-09-972-467-2  
; Sequence 2, Application US/09972467  
; Patent No. US20020090373A1  
; GENERAL INFORMATION:  
; APPLICANT: PFIZER INC.  
; TITLE OF INVENTION: ADAMTS POLYPEPTIDES, NUCLEIC ACIDS ENCODING THEM, AND  
; FILE OF INVENTION: USES THEREOF  
; FILE REFERENCE: PC10850A  
; CURRENT APPLICATION NUMBER: US/09/972,467  
; CURRENT FILING DATE: 2001-10-05  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 1629  
; TYPE: PRT  
; ORGANISM: Human  
US-09-972-467-2  
  
Query Match 7.3%; Score 597.5; DB 9; Length 1629;  
Best Local Similarity 21.3%; Pred. No. 1.3e-37;  
Matches 225; Conservative 102; Mismatches 328; Indels 401; Gaps 50;  
  
QY 201 WSS---WSPCKTCSRSGSLLPGRFSRNRVKNHMAIGGKCEPILLEKEACIVGELL--- 254  
DB 881 WNSGHPQACSKPCQ-----GERKR-----KLVCTRSDQLTVS 914  
QY 255 -QCPRYSNRTSEWKEQVSLLEQQDPHWHVT-----GVCVGGIQTREVVYCAQSVFAA 308  
DB 915 DQCDRLFPQGHITPCGTDC-----DLRWHASRSECSAQCGLYRTLDIYCAK----- 964  
QY 309 AALRAKEVSRPVEKALCVGAPLPSQLCNIPGCTDCIVSSW--SAWGLCIHENCHEPQK 366  
DB 965 -YSLDCKTEKVDGFC--SSHPPSN--REKCSGECNTGWRYSANTEC----- 1009  
QY 367 KGFTROHRLVMESTGPAGHCPHLVSPCEDPMCVRWLASSEGICFPDGHGKGLGHRILK 426  
DB 1010 -----SKSCDGGTQRRR 1021  
QY 427 AVCONDRGEDVSGSLCPVPPPPERKSC-EIPCRMDCVLSEWT--EWSSCSQSCSNKNSDG 483  
DB 1022 AICVNTNRNDVDDSKTHQEKVTIQCSEFFC-----POMKSGDMSCLVTC-----G 1069  
QY 484 KQTSRRTILAGE--GGKPCFPFSQLQHLRCLNDHSCMLHWETSPWGPCSEDTLVTA 540  
DB 1070 KGHKRVQVWCQFGEEDRLNDRMCDPETKPTSMQTCQPECAS--WQAGPWGQCS----- 1120  
QY 541 LNATIGNGATCGVGIOTRREVC-VKSHVQVMTKRCPSDSTRPETVRPCFLPKCKDC-- 597  
DB 1121 -----VTCGGYQLRAVVKCIIGTYMSVVDDNDCNAATRTDTQCELP---SCHP 1167  
QY 598 -----IVTAFSEWTPCPKRCQAGNATVKQSRVRIIIEAANGQECPCPT 641  
DB 1168 PPAAPETRRSTYSAPRTQWRFGSWTPCSATCGK-----TRVRYVSCRDENG-----S 1215  
QY 642 LYBEREC-----EDVSLCPVYKWKPKWSPCLVPSVWQGITGSEACGKLOTR 692  
DB 1216 VADESACATLPRPVAKKECVSTPCGWKALDWSSC-----SVTCQGRATR 1261  
QY 693 AVSCISDNRSNAEMCECLKQTNMGVPLLVQECTVPCREDCTFTAWSKFTPCSTNCEATKSR 752  
DB 1262 QVMCVNTSDHVIDRESC--DQYIPETDQDCSM-----SPC----- 1295  
QY 753 RRLQTKSRKKEKQDSLD---YPLVETELCPDEFTISQPY-----GNWSDCILPEG 801  
DB 1296 -----PORTPSGLAQHPFQNEEDYRPSASPSRTHVLGNGQWRTPGPMGAC----- 1340  
QY 802 RREPHRGLRVQADKECEGLRFRVACSKNG-----RPVDPSPCSSSGVIOEK 851  
DB 1341 -----SSTCAGGSQRRVVVQDENGYTANDVERIKPDEORACESG----- 1381  
QY 852 CVTFCPFDCKLSDWS--SWGSCSSSCGIGVIRSRKWLKEKPYNGRCPKPLDLKNQVHEA 909



```
Db 531 EEPSPFPEAMSACTVTCGVTQVIRVRCQVLLSFSQSVADLPIDECEGPKPA-----SOR 585
2Y 906 VHEAVPCYSCNQ-----YSWVVEHWSCKINNELSLRSCGGGTOS 946
Db 586 ACVAGPCSGEIPFNPDETDLGEGGLQDFEYDWEYEGFTK-----SESCGGVQOE 638
2Y 947 RKIRCVNTADGEGAVDSNLCNDEIPPE--TQCSLMCPNECYMSEWGLWSKQSCDPH 1005
Db 639 AVVSCLNKQTR--PAENLCVTSRRPPQLKSCNL-----DPCPA-----677
2Y 1006 TMQERTHLLRPSLNSRTCAEDSQVQPCLLNENCFOFQYNLTWSTCOLSENAPCGQVR 1065
Db 678 -----RWEIGKWSPCSLT-----CGVGLQ 696
QY 1066 TRLLSC-----VCSGDKPVSMDQCEQHNLKRPQMSIPCLVECVVNCQ 1108
Db 697 TRDVFCSHLLSRENNETVILADELCRQKPESTVQACNR-----FNCP 738
QY 1109 LSGHTA-WTECSQTCGHG-----GRMSRTRFIIMPTQ--GEGRPCPTELTQKTC 1156
Db 739 PAMYPACQPCSRTCGGGVQKREVLCKQRMADGSFLELPETFCASQACQACKKDDCP 798
QY 1157 VTPCYSVLGNWSACKLEGGDCGEGVQIRSLSC--MVHSGSIHAAGRVEDALCGEMPFQ 1214
Db 799 ----SEWLLSDWTECST---SCGEGTQTRSAICRKMKTG-ISTV---VNSTLCPLPFS 847
QY 1215 DSILKQLCSVPC 1226
Db 848 SSIR-----PC 853

RESULT 15
US-10-044-807-2
; Sequence 2, Application US/10044807
; Publication No. US20020165187A1
; GENERAL INFORMATION:
; APPLICANT: Yu, Xuanchuan
; TITLE OF INVENTION: Turner, C. Alexander Jr.
; FILE REFERENCE: LEX-0298-USA
; CURRENT APPLICATION NUMBER: US/10/044,807
; CURRENT FILING DATE: 2002-01-11
; PRIOR APPLICATION NUMBER: US 60/261,684
; PRIOR FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1762
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-044-807-2

Query Match 6.0%; Score 496; DB 13; Length 1762;
Best Local Similarity 19.8%; Pred. No. 1.4e-29;
Matches 228; Conservative 127; Mismatches 324; Indels 474; Gaps 53;

QY 201 WSSWSPCKTCRSG---SLLPGFSSRS---RNVKHMAIGGKSCPEL---LEKXACIVEG 251
Db 39 WGPWSECSRTCCGGASYSRLRCLSSKCEGRIYRTC-SNVDCPPRAGDFRAQQCSAHN 97
QY 252 ELLQCCPRYSWRTSEWKECVSLLEQQDPPHHVTGPVCGGIGTREVYCAQSVPAAL 311
Db 98 DVKHGGQFYEW-----LPVNDP-----DNFCSLKC 123
QY 312 RAKEVSRPVEKALCVGAPLPSQJCNIPCTDCTVSSWSAWGLCIHNCHEPQCKGFR 371
Db 124 QAKGTTLVELA-----PKVLDGTRCYTESL-----DMCISGLC-----157
QY 372 RQRVLMESTGAGCHPLVESVCEPDMCMYRWLASRGICFPDHGKGLGHLKAVCON 431
Db 158 -----QIVG-----CDHQLGSTVKED-----NCGVNCNGDSTCLRVGRQYKSLSA 198
```

```
QY 432 DRGEDV-----SGSLCPVPVPPPERKSCFIPCMDCVLSEWTESSSCSCSNKNSDGK 484
Db 199 TKSDTTVAIPYSGSRHRLVLKGPDLHLYLETK-TLOGTKGENSLSTGTFLVDNSSVDFQ 257
QY 485 QTRSDTILALAGEGKCPSPSOALOEHRLCNDHSCML-----HWETSPWGPCSEDT 536
Db 258 KFPDKELIRWAG-----PLTADFIVKIRNSGASDTVOFIYQPIIHRWRTDFFPCS---310
QY 537 LVTALNATIGNWGECATCGVGIQTRRVFCVKSHVQVMTKR-C---PDSTRPETVRPCLP 592
Db 311 -----ATCGGGYQLTSAECYDLRSNRVADQVCHYYPENIKP---KPKLQE 353
QY 593 CKDCIYTAISEWTECPRMQAGNATVQKSYRIIIQEAANG-GOECPDLYBERECEDV 651
Db 354 CNLD-----PCP-----ASDGYKQIMPDIYH-----375
QY 652 SLCPYRWKPKOKSPCILVPESVWQGITSGSEACGKGLQTRAVSCISDDNR-----SAEMM 707
Db 376 ---PLPRWEATPWATAC-----SSCGGIGQSRVSCVEEDIQGHVTSVEEW 418
QY 708 ECLQQTNGMPLLVQECTVPCREDCTFTAWSKFTPCSTNCEATKSRRLQITGKRKKEKCQ 767
Db 419 KCM-YTPMPPIA-----QPCN-----433
QY 768 DSDLYPLVETELCPDCEFIQPYGNWSDCILPEGRREPHRLRVQADSKECCEGLRPRAV 827
Db 434 -----IFDQPKWLAQ---EMSPCTV-----TCGQGLRYRVV 461
QY 828 ACSDKNGRPVDPSPFCSSSGYIOEKCVP-----CPFDCKLSDW-----865
Db 462 LCIDHRGMHTGCGSPKTKPHIKEECIVPTPCYKPEKLPVEAKL-PWFQQAQAELEGA 520
QY 866 -----SSWGSCTSSCGIGVRIR-----SKWLKEKPNY---GGRPCPKLDLN 904
Db 521 SEEPSFPEAMSACTVTCGVTQVIRVRCQVLLSFSQSVADLPIDECEGPKPA-----SQ 575
QY 905 QVHEAVPCYSCNQ-----YSWVVEHWSCKINNELSLRSCGGGTQ 945
Db 576 RACYAGPCSGEIPFNPDETDLGEGGLQDFEYDWEYEGFTK-----SESCGGVQ 628
QY 946 SRKIRCVNTADGEGAVDSNLCNDEIPPE--TQCSLMCPNECYMSEWGLWSKQSCDP 1004
Db 629 EAVVSCLNKQTR--PAENLCVTSRRPPQLKSCNL-----DPCPA-----668
QY 1005 HTMQRTTRHLLRPSLNSRTCAEDSQVQPCLLNENCFOFQYNLTWSTCOLSENAPCGV 1064
Db 669 -----RWEIGKWSPCSLT-----CGVGL 686
QY 1065 RTLELSC-----VCSGDKPVSMDQCEQHNLKRPQMSIPCLVECVVNC 1107
Db 687 QTRDVFCSHLLSRENNETVILADELCRQKPESTVQACNR-----FNC 728
QY 1108 QLSGWTA-WTECSQTCGHG-----GRMSRTRFIIMPTQ--GEGRPCPTELTQKTC 1155
Db 729 PPAWYPAQWPCSRTCGGGVQKREVLCKQRMADGSFLELPETFCASQACQACKKDDC 788
QY 1156 PTPCYSVLGNWSACKLEGGDCGEGVQIRSLSC--MVHSGSIHAAGRVEDALCGEMPF 1213
Db 789 P-----SEWLLSDWTECST---SCGEGTQTRSAICRKMKTG-----LSTVNVNSTLCPLP 837
QY 1214 QDSILKQLCSVPC 1226
Db 838 SSIR-----PC 844

Search completed: February 24, 2004, 01:30:05
Job time : 619.973 secs
```



GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

M nucleic - nucleic search, using sw model

Run on: February 23, 2004, 10:50:55 ; Search time 1592.77 Seconds  
(without alignments)  
11538.043 Million cell updates/sec

Title: US-10-022-710-5\_COPY\_1\_424

Perfect score: 424  
Sequence: 1 agtttccaaagaaact.....gcatgactgcagcacogga 424

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*

1: gb\_ba:\*

2: gb\_hg:\*

3: gb\_in:\*

4: gb\_om:\*

5: gb\_ov:\*

6: gb\_pat:\*

7: gb\_ph:\*

8: gb\_pl:\*

9: gb\_pr:\*

10: gb\_ro:\*

11: gb\_sta:\*

12: gb\_sy:\*

13: gb\_un:\*

14: gb\_vt:\*

15: em\_ba:\*

16: em\_fun:\*

17: em\_hum:\*

18: em\_in:\*

19: em\_mu:\*

20: em\_om:\*

21: em\_or:\*

22: em\_ov:\*

23: em\_pat:\*

24: em\_ph:\*

25: em\_pl:\*

26: em\_ro:\*

27: em\_sta:\*

28: em\_un:\*

29: em\_vt:\*

30: em\_hg\_hum:\*

31: em\_hg\_inv:\*

32: em\_hg\_other:\*

33: em\_hg\_mus:\*

34: em\_hg\_pin:\*

35: em\_hg\_rod:\*

36: em\_hg\_mam:\*

37: em\_hg\_vrt:\*

38: em\_sv:\*

39: em\_hggo\_hum:\*

40: em\_hggo\_mus:\*

41: em\_hggo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	294.4	69.4	6112	10	BC063250 Mus muscu
C 2	290.4	68.5	119205	9	AC011231 Homo sapi
C 3	290.4	68.5	167462	2	AC027600 Homo sapi
C 4	209	49.3	5724	9	AB051466 Homo sapi
C 5	200	47.2	224789	2	AC116111 Rattus no
C 6	195.2	46.0	259236	2	AC123741 Mus muscu
C 7	141	33.3	164366	2	AC079808 Homo sapi
C 8	141	33.3	180602	9	AC010873 Homo sapi
C 9	110.6	26.1	234528	2	AC096222 Rattus no
C 10	104.2	24.6	175131	2	AC124806 Mus muscu
C 11	78.6	18.5	3053	6	AX747433 Sequence
C 12	78.6	18.5	3053	9	AK092252 Homo sapi
C 13	76	17.9	6378	6	AR338841 Sequence
C 14	75.4	17.8	6373	6	AX079870 Sequence
C 15	75.4	17.8	112456	9	AC004614 Homo sapi
C 16	75.4	17.8	170419	9	AC146265 Pan trogl
C 17	75.4	17.8	184549	2	AC068575 Homo sapi
C 18	73.8	17.4	187347	2	AC135033 Rattus no
C 19	73.8	17.4	234459	2	AC127064 Rattus no
C 20	70.6	16.7	172164	2	AC115892 Mus muscu
C 21	36.6	8.6	275073	2	AC115176 Rattus no
C 22	36.4	8.6	238386	2	AC109763 Rattus no
C 23	36.4	8.6	345974	2	AC110656 Rattus no
C 24	36.2	8.5	4320	10	AK129133 Mus muscu
C 25	36.2	8.5	8456	10	AB072381 Mus muscu
C 26	36.2	8.5	215581	2	AC106161 Rattus no
C 27	36.2	8.5	230955	2	AC095647 Rattus no
C 28	36.2	8.5	235977	2	AC132507 Rattus no
C 29	35.8	8.4	1629	6	AX167247 Sequence
C 30	35.8	8.4	3460	10	BC031804 Mus muscu
C 31	35.8	8.4	166743	10	AC127226 Mus muscu
C 32	35.8	8.4	182921	10	AC124761 Mus muscu
C 33	35.8	8.4	183707	10	AC124384 Mus muscu
C 34	35.6	8.4	163209	9	AC093419 Homo sapi
C 35	35.4	8.3	132169	2	AC026814 Mus muscu
C 36	35.4	8.3	175144	2	AC068502 Mus muscu
C 37	35.4	8.3	209857	10	AL583883 Mouse DNA
C 38	35.2	8.3	124831	2	AC121270 Mus muscu
C 39	35.2	8.3	214452	2	AC121752 Rattus no
C 40	35	8.3	11505	1	AE001936 Deinococc
C 41	35	8.3	124256	2	AC124151 Oryza sat
C 42	35	8.3	166300	9	AC018884 Homo sapi
C 43	35	8.3	179692	2	BX539346 Danio rer
C 44	35	8.3	185773	2	AC027528 Homo sapi
C 45	35	8.3	206214	2	AC079859 Homo sapi

# ALIGNMENTS

RESULT 1  
BC063250  
LOCUS  
DEFINITION  
BC063250 Mus musculus RIKEN cDNA D130067103 gene, mRNA (CDNA clone MGC:166597 IMAGE:6410178), complete cds.  
ACCESSION  
BC063250  
VERSION  
BC063250.1 GI:38649124  
KEYWORDS  
MGC.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
REFERENCE  
1 (bases 1 to 6112)  
Srausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,  
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,  
linear ROD 16-DEC-2003



RESULT 2  
AC011231/c

LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AC011231 119205 bp DNA linear PRI 07-NOV-2001  
Homo sapiens BAC clone RP11-13K3 from 2, complete sequence.  
AC011231.7 GI:14190752  
HTG.  
Homo sapiens (human)

REFERENCE  
AUTHORS

TITLE  
JOURNAL  
MEDLINE  
PUBMED

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 119205)  
Sulston, J.E. and Waterston, R.  
Toward a complete human genome sequence  
Genome Res. 8 (11), 1097-1108 (1998)  
99063792  
9847074

REFERENCE  
AUTHORS

TITLE  
JOURNAL  
MEDLINE  
PUBMED

2 (bases 1 to 119205)  
Harris, A. and Maupin, R.  
The sequence of Homo sapiens BAC clone RP11-13K3  
Unpublished  
3 (bases 1 to 119205)  
Waterston, R.H.  
Direct Submission  
Submitted (04-OCT-1999) Genome Sequencing Center, Washington  
University School of Medicine, 444 Forest Park Parkway, St. Louis,  
MO 63108, USA

REFERENCE  
AUTHORS

TITLE  
JOURNAL  
MEDLINE  
PUBMED

4 (bases 1 to 119205)  
Waterston, R.H.  
Direct Submission  
Submitted (23-MAY-2001) Genome Sequencing Center, Washington  
University School of Medicine, 444 Forest Park Parkway, St. Louis,  
MO 63108, USA

REFERENCE  
AUTHORS

TITLE  
JOURNAL  
MEDLINE  
PUBMED

5 (bases 1 to 119205)  
Waterston, R.H.  
Direct Submission  
Submitted (07-NOV-2001) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On May 23, 2001 this sequence version replaced gi13431080.

COMMENT

----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@wustl.edu](mailto:sapiens@wustl.edu)  
----- Summary Statistics  
Center project name: H\_NH0013K03  
-----

NOTICE: This sequence may not represent the entire insert of this  
clone. It may be shorter because we only sequence overlapping  
clone sections once, or longer because we provide a small overlap  
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate  
chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by sequence  
from more than one subclone; and the assembly was confirmed by  
restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D.  
McPherson, Department of Genetics, Washington University, St. Louis  
MO. For additional information about the map position of this  
sequence, see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RP11-11 human BAC library was made from the blood of one male  
donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E.,  
Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved

approach for construction of bacterial artificial chromosome  
libraries. Genomics 51:1-8. The clone may be obtained either from  
Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong  
and coworkers at the Roswell Park Cancer Institute  
(<http://bactpac.med.buffalo.edu>)  
VECTOR: pBAC3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-398K16, 200 bp overlap; the  
clone sequenced to the right is RP11-112N16. Actual start of this  
clone is at base position 138700 of RP11-398K16; actual end is at  
base position 7193 of RP11-112N16.

#### FEATURES

Location/Qualifiers

source

1. 119205  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/map="2"

/clone="RP11-13K3"

/clone\_lib="RPC1-11"

repeat\_region

2. 164

repeat\_region

/rpt\_family="L2"

repeat\_region

7728. 7785

repeat\_region

/rpt\_family="L2"

repeat\_region

7805. 8116

repeat\_region

/rpt\_family="Alu"

repeat\_region

8482. 8544

repeat\_region

/rpt\_family="L2"

repeat\_region

8671. 8748

repeat\_region

/rpt\_family="Mariner"

repeat\_region

10478. 10853

repeat\_region

/rpt\_family="L2"

repeat\_region

11336. 11375

repeat\_region

/rpt\_family="(TG)n"

repeat\_region

11737. 11771

repeat\_region

/rpt\_family="AT-rich"

repeat\_region

12229. 12423

repeat\_region

/rpt\_family="MIR"

repeat\_region

13139. 13436

repeat\_region

/rpt\_family="Alu"

repeat\_region

13554. 13694

repeat\_region

/rpt\_family="MIR"

repeat\_region

13704. 13718

repeat\_region

/rpt\_family="MIR"

repeat\_region

13719. 13765

repeat\_region

/rpt\_family="L2"

repeat\_region

13766. 13892

repeat\_region

/rpt\_family="MIR"

repeat\_region

14019. 14272

repeat\_region

/notes="similar to EST BE299433 (NID:G9183181)"

repeat\_region

14575. 14855

repeat\_region

/rpt\_family="Alu"

repeat\_region

14903. 15059

repeat\_region

/rpt\_family="MIR"

repeat\_region

15247. 15381

repeat\_region

/rpt\_family="L1"

repeat\_region

15621. 15826

repeat\_region

/rpt\_family="MER1\_type"

repeat\_region

16211. 16507

repeat\_region

/rpt\_family="Alu"

repeat\_region

16695. 16988

repeat\_region

/rpt\_family="Alu"

repeat\_region

17012. 17049

repeat\_region

/rpt\_family="AT-rich"

repeat\_region

17297. 17621

repeat\_region

/rpt\_family="MaLR"

repeat\_region

17622. 17653

repeat\_region

/rpt\_family="(TAAA)n"

repeat\_region

17654. 17715

repeat\_region

/rpt\_family="MaLR"

repeat\_region

19019. 19049

repeat\_region

/rpt\_family="AT-rich"

repeat\_region

19050. 19108

```

repeat_region      /rpt_family="L1"
19109..19386
/rpt_family="Alu"
repeat_region      19387..19705
/rpt_family="L1"
repeat_region      19706..19870
/rpt_family="(TA)n"
repeat_region      19890..20882
/rpt_family="L1"
repeat_region      20883..21186
/rpt_family="Alu"
repeat_region      21187..21558
/rpt_family="L1"
repeat_region      21559..21583
/rpt_family="(TTT)G)n"
repeat_region      21584..21872
/rpt_family="L1"
repeat_region      21950..21984
/rpt_family="AT-rich"
repeat_region      22011..22261
/rpt_family="L1"
repeat_region      22474..22706
/rpt_family="WALR"
repeat_region      22814..23079
/rpt_family="Alu"
repeat_region      23190..23461
/rpt_family="L1"
repeat_region      23518..24159
/rpt_family="L1"
repeat_region      24270..24652
/rpt_family="WALR"
repeat_region      24696..24959
/rpt_family="L1"
repeat_region      26424..26590
/rpt_family="MIR"
repeat_region      27884..28177
/rpt_family="L1"
repeat_region      28632..28930
/rpt_family="Alu"
repeat_region      28940..28990
/rpt_family="(CATATA)n"
repeat_region      29009..29317
/rpt_family="Alu"
repeat_region      29554..29581
/rpt_family="L1"
repeat_region      29982..30099
/rpt_family="Alu"
repeat_region      30100..30583
/rpt_family="L1"
repeat_region      30826..32314
/rpt_family="L1"
repeat_region      38581..38608
/rpt_family="AT-rich"
repeat_region      38609..38984
/rpt_family="L1"
repeat_region      39073..39148
/rpt_family="L1"
repeat_region      39214..39340

Query Match      68.5%; Score 290.4; DB 9; Length 119205;
Best Local Similarity 86.3%; Pred. No. 9.9e-82;
Matches 321; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 53 GGAAGCTCTTCTATGTTCTCTCTTGTGTCCTCCATGCAAGCTCATTTGGAAGGCAAAA 112
DB 52812 GGAAGTGTGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 52753
QY 113 AGGATAATCAGTTTCATCTGGAACACAGTCCGTGGGAGGTACAGGAGACTGTGTC 172
DB 52752 TTAACATCTTGTGTTTCTGCTGTAGTCCGTGGGAGGTACAGGAGACTGTGTC 52693
QY 173 CCGGAGGAGTCCAGAGTCCGGCAGTGTGTTTTCATGTGTTGACGGTGGACAGTCACC 232

```

```

Db 52692 CCGGAGGAGTCCAGAGTCCGGCAGTGTGTTTTCATGTGTTGACGGTGGACAGTCACC 52633
QY 233 TGTCTAACTGTGGTGAGAGCAACAGGCTCCAAAGGAAAGAGTGTGTTTCCGAGTTTGTG 292
Db 52632 TGTCTAACTGTGGTGAGAGCAACAGGCTCCAAAGGAAAGAGTGTGTTTCCGAGTTTGTG 52573
QY 293 ACTGGCAGAGTGAACCTCTTTTCAGTGGGAGGTTTCTGACTGGCACCACCTGTGTGTTTC 352
Db 52572 ACTGGCAGAGTGAACCTCTTTTCAGTGGGAGGTTTCTGACTGGCACCACCTGTGTGTTTC 52513
QY 353 CTTAGCTCGGGGTGAGTCAAGCTCGGACTGAGAGTGTGTTGACCGGCTCAGCATGGAC 412
Db 52512 CTTAGCTCGGGGTGAGTCAAGCTCGGACTGAGAGTGTGTTGACCGGCTCAGCATGGAC 52453
QY 413 TGCAGCACCGGA 424
Db 52452 TGCAGCACCGGA 52441

RESULT 3
AC027600/c
LOCUS
DEFINITION      Homo sapiens chromosome 2 clone RP11-41418 map 2, WORKING DRAFT
SEQUENCE, 22 unordered pieces.
ACCESSION      AC027600
VERSION        AC027600.2 GI:8318467
KEYWORDS       HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE         Homo sapiens
ORGANISM       Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 167462)
AUTHORS        Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE          Homo sapiens chromosome 2, clone RP11-41418
JOURNAL        Unpublished
REFERENCE      2 (bases 1 to 167462)
AUTHORS        Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bada,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kamm,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczyk,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marcuis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Olivaz,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
JOURNAL        Direct Submission
COMMENT        Submitted (30-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 7, 2000 this sequence version replaced gi:7342347.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
Smith, A.F.A. & Green, P. (1996-1997)
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L9157
Center clone name: 414_I_8
----- Summary Statistics

```

Sequencing vector: M13; M77815; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 156843 bases at least Q40  
 Consensus quality: 161829 bases at least Q30  
 Consensus quality: 163971 bases at least Q20  
 Insert size: 170000; agarose-fp  
 Quality coverage: 4.2 in Q20 bases; agarose-fp  
 Quality coverage: 4.3 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 22 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1293: contig of 1293 bp in length  
 1294 1393: gap of 100 bp  
 1394 3228: contig of 1835 bp in length  
 3229 3329: gap of 100 bp  
 3329 5140: contig of 1812 bp in length  
 5141 5240: gap of 100 bp  
 5241 8017: contig of 2777 bp in length  
 8018 8117: gap of 100 bp  
 8118 10466: contig of 2349 bp in length  
 10467 10566: gap of 100 bp  
 10567 13763: contig of 3197 bp in length  
 13764 13863: gap of 100 bp  
 13864 15896: contig of 2033 bp in length  
 15897 15996: gap of 100 bp  
 15997 19987: contig of 3991 bp in length  
 19988 20087: gap of 100 bp  
 20088 24063: contig of 3976 bp in length  
 24064 24163: gap of 100 bp  
 24164 26824: contig of 2661 bp in length  
 26825 26924: gap of 100 bp  
 26925 30928: contig of 4004 bp in length  
 30929 31028: gap of 100 bp  
 31029 34504: contig of 3476 bp in length  
 34505 34604: gap of 100 bp  
 34605 40409: contig of 5805 bp in length  
 40410 40509: gap of 100 bp  
 40510 46642: contig of 6133 bp in length  
 46643 46742: gap of 100 bp  
 46743 55317: contig of 8575 bp in length  
 55318 55417: gap of 100 bp  
 55418 64038: contig of 8621 bp in length  
 64039 64138: gap of 100 bp  
 64139 71781: contig of 7643 bp in length  
 71782 71881: gap of 100 bp  
 71882 81417: contig of 9536 bp in length  
 81418 81517: gap of 100 bp  
 81518 102057: contig of 20540 bp in length  
 102058 102157: gap of 100 bp  
 102158 121599: contig of 19442 bp in length  
 121600 121699: gap of 100 bp  
 121700 142729: contig of 21030 bp in length  
 142730 142830: gap of 100 bp  
 142830 167462: contig of 24633 bp in length.

# FEATURES

Location/Qualifiers  
 1. .167462  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="2"  
 /map="2"  
 /clone="RP11-41418"  
 /clone\_lib="RPC1-11 Human Male BAC"  
 1. .1293  
 /note="assembly\_fragment"  
 misc\_feature

misc\_feature 1394. .3228  
 /note="assembly\_fragment"  
 misc\_feature 3329. .5140  
 /note="assembly\_fragment"  
 misc\_feature 5241. .8017  
 /note="assembly\_fragment"  
 misc\_feature 8118. .10466  
 /note="assembly\_fragment"  
 misc\_feature 10567. .13763  
 /note="assembly\_fragment"  
 misc\_feature 13864. .15896  
 /note="assembly\_fragment"  
 misc\_feature 15997. .19987  
 /note="assembly\_fragment"  
 misc\_feature 20088. .24063  
 /note="assembly\_fragment"  
 misc\_feature 24164. .26824  
 /note="assembly\_fragment"  
 misc\_feature 26925. .30928  
 /note="assembly\_fragment"  
 misc\_feature 31029. .34504  
 /note="assembly\_fragment"  
 misc\_feature 34605. .40409  
 /note="assembly\_fragment"  
 misc\_feature 40510. .46642  
 /note="assembly\_fragment"  
 misc\_feature 46743. .55317  
 /note="assembly\_fragment"  
 misc\_feature 55418. .64038  
 /note="assembly\_fragment"  
 misc\_feature 64139. .71781  
 /note="assembly\_fragment"  
 misc\_feature 71882. .81417  
 /note="assembly\_fragment"  
 misc\_feature 81518. .102057  
 /note="assembly\_fragment"  
 misc\_feature 102158. .121599  
 /note="assembly\_fragment"  
 misc\_feature 121700. .142729  
 /note="assembly\_fragment"  
 misc\_feature 142830. .167462  
 /note="assembly\_fragment"

## ORIGIN

Query Match 68.5%; Score 290.4; DB 2; Length 167462;  
 Best Local Similarity 86.3%; Pred. No. 1e-81;  
 Matches 321; Conservative 0; Mismatches 51; Indels 0; Gaps 0;  
 Qy 53 GGAGCTCTTCTATTGCTTCTCTCTGCTGCCATGCAGCTCATTTGGAGGCAAAA 112  
 Db 145485 GGAAGTGTGTTAATGACTTCTACTTACTCTGAATGCTGAATGTAATCTCTGAGTAA 145426  
 Qy 113 AGGATAATCAGTTTCATCTGGAACACAGGTCCTGGGGAAGGTGTACAGGAGACTGTGGTC 172  
 Db 145425 TTAACATCTTGTGTTTCTGCTGCTAGTCCGTTGGGGAAGGTGTACAGGAGACTGTGGTC 145366  
 Qy 173 CCGAGAGTCCAGAGTCGGGCGAGTGTGTTTCTATGTTGCGGTGGCAAGTCACC 232  
 Db 145365 CCGAGAGTCCAGAGTCGGGCGAGTGTGTTTCTATGTTGCGGTGGCAAGTCACC 145306  
 Qy 233 TGCTTAATCTGTTGAGAGCAACAGGCTCCAAAGGAAAGTGTGTTTCCGAGTTGTG 292  
 Db 145305 TGCTTAATCTGTTGAGAGCAACAGGCTCCAAAGGAAAGTGTGTTTCCGAGTTGTG 145246  
 Qy 293 ACTGGCAGTGACCTCTTTTCAGTGGGAGGTTTCTGACTGGCAGCACTGTGCTGTGTC 352  
 Db 145245 ACTGGCAGTGACCTCTTTTCAGTGGGAGGTTTCTGACTGGCAGCACTGTGCTGTGTC 145186  
 Qy 353 CTTACGCTCGCGGTGAAGTCAAGCTCGGACTGCAGAGTGTGTGACGGCTCAGCATGGAC 412

|||||  
145185 CTTACGCTCGCGTGAAGTCAAGCTCGACTGAGAGTGTGTGACGGCTCAGATGGAC 145126  
|||||  
QY 413 TGCAGCACCGGA 424  
|||||  
Db 145125 TGCAGCACCGGA 145114  
|||||  
RESULT 4  
AB051466 5724 bp mRNA linear PRI 07-FEB-2001  
LOCUS Homo sapiens mRNA for KIAA1679 protein, partial cds.  
DEFINITION AB051466  
ACCESSION AB051466  
VERSION AB051466.1 GI:12697902  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (sites)  
Nagase, T., Kikuno, R., Hattori, A., Kondo, Y., Okumura, K. and Ohara, O.  
Prediction of the coding sequences of unidentified human genes.  
XIX. The complete sequences of 100 new cDNA clones from brain which  
code for large proteins in vitro  
DNA Res. 7 (6), 347-355 (2000)  
21082932  
11214970  
2 (bases 1 to 5724)  
Ohara, O., Nagase, T. and Kikuno, R.  
Direct Submission  
Submitted (22-NOV-2000) Osamu Ohara, Kazusa DNA Research Institute,  
Department of Human Gene Research; 1532-3, Yana, Kisarazu, Chiba  
292-0812, Japan (E-mail: cdnainfo@kazusa.or.jp).  
URL: <http://www.kazusa.or.jp/huge>, Tel: 81-438-52-3913,  
Fax: 81-438-52-3914  
Location/Qualifiers  
1. 5724  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="fn22954"  
/note="vector:pBluescriptII SK plus"  
1. 5724  
/gene="KIAA1679"  
/note="KIAA1679"  
/note="Start codon is not identified."  
/codon\_start=2  
/product="KIAA1679 protein"  
/protein\_id="BAB21770.1"  
/db\_xref="GI:12697903"  
/translation="GWTSHLSNCGESNRPKERSCFVCDWHSDFOWEVDWHHCVL  
VPYARGEVKPTAEVTAQHGLHMRVRCIOKLNRTVANEICEHFALPPTQACLI  
PCPRDVCVSEFLPWSNCSKGKQLOHRTAVIAPLFGLOCPNLTESRACDPLSC  
PLGEEVTFSLKVPWPKSLPLHKEINPSEGTVDNPSNSRNVTFKHQSYXAHHS  
KSWAIEGTQKVCSTNSDQAWLSLCQDSFPLTVQSCIMPCKOCETQSNWSWSPC  
SKTRSGSLFGRSRNRVHMAIGGCKPELLEKEACEICVIGELQQCPRYSWRTS  
EWKCSLLLEQDPHWHVTGVCVGGIQTREYVCAQSPAAALAEKESVPEKA  
LCVGPAPLPSQCLNPGSTDCIVSSWAGWLCIHENHEPQKGFRTQRHLMEST  
GPAGCHPLFVSPCEPDMCVLASEGICFPDHGKGLGHRILKAVCONDRGDEVSG  
SLCPVPPPEPKSCIEPCMDCVLSEWTESSCSQSNKNSDQKPSRPTILALAGE  
CGKCPSPQALQHLRNLNDHSCMLHNETSPWGPCSDETLVLTALNATIGNGATCGV  
GIQTRRVKSHVGVQWTKHSCQDSTPRTPEVTRPFCFLKDCDCTVATSEFWPTCFMVCQ  
AGNATVKOSRVRIIIOEANGGQCEPDLTYEERECEDVSLCPVVRWPKWSKPLVP  
ESVAGTIGTSSACGKGLQTVASCIIDNRSKAEEMECLOTGMLPVLLOVECTVPCRE  
DCTTANSKFTPCSTNEATRSRQLTGKSKKEKQDSDPLVLVTELCPCDEFTS  
QPYGWSDCIPEGRPHGLRVADSKCEGCEFRVACDKNGRPVDPSPCS  
GVIOEKVCIFCFCKLSDSSWGS CSSCGIGVIRISKLKEPYNGRCPCKLDLK  
NOAQVHEAVPCYSCNQSWYVHWSCKINNELRSRCGGGTQSRKIRCVNTADGEG  
GAVSNLCAEDQVQPCNLNCFQYNLTWSTQLSENAPCGQVTRLLSCVCSDF  
SLNRSKTCBDSQVQPCNLNCFQYNLTWSTQLSENAPCGQVTRLLSCVCSDF  
GKPSVMDQCEQHLEKQRMISICPLVCEVNCQSLGTAWTETCSQTCGCHGRSRTFR  
IIMPTQSGRCPCTELTQKTCPTVPCYVWLGWNSACKLEGSGDGBGVQIRSLSCW

|||||  
HSGSISHAAGRVEDALCGEMPFPQDSILKQLCSVPCGDCHLTSEWSTCELTICIDGR  
SFEVVGSRSRRTIIIOFENQDSCPOQVLETRPTCTGCKGCVHYTWKASLWNNERTYV  
CORSQDGNVNTGGCSPQARPAIROCIIPACKRPSYCTGGVCGEKGTEIMRNGFL  
DYCKWPGSEDKADVKNLGKRNFKNSKIHDIKFGWSLOPLDPRGKVIWVYVSGG  
APLIMIFLIFTSYLVCKKPKPHQSTPQPKRELITLAYDGLDM"  
|||||  
QY 216 CGGGTGGACAGTACCTGTCTTAAGTGGTGAGAGCAACAGGCTCCAAAGGAAG 275  
|||||  
Db 1 CGGGTGGACAGTACCTGTCTTAAGTGGTGAGAGCAACAGGCTCCAAAGGAAG 60  
|||||  
QY 276 TTGTTTCCGAGTTTGTGACTGGCAGTGCCTCTTTTCAGTGGGAGGTTTCTGACTGGCA 335  
|||||  
Db 61 TTGTTTCCGAGTTTGTGACTGGCAGTGCCTCTTTTCAGTGGGAGGTTTCTGACTGGCA 120  
|||||  
QY 336 CCACGTGTGCTGTGTTCTTACGTCGGGTGAGTGAAGCTCGGACTCGGACTCGAGAGTGTGT 395  
|||||  
Db 121 CCACGTGTGCTGTGTTCTTACGTCGGGTGAGTGAAGCTCGGACTCGGACTCGAGAGTGTGT 180  
|||||  
QY 396 GACGGCTCAGCATGACTGCAGCACCGCA 424  
|||||  
Db 181 GACGGCTCAGCATGACTGCAGCACCGCA 209  
|||||  
RESULT 5  
AC111611 224789 bp DNA linear HTG 13-NOV-2002  
LOCUS Rattus norvegicus clone CH230-237P17, \*\*\* SEQUENCING IN PROGRESS  
DEFINITION \*\*\*  
AC111611  
AC111611.4 GI:24942115  
VERSION HTG; HTGS PHASE2; HTGS\_DRAFT; HTGS\_ENRICHED.  
KEYWORDS Rattus norvegicus (Norway rat)  
SOURCE Rattus norvegicus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 224789)  
Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,  
Allen, C., Allen, R., Albrooks, S., Amin, A., Anguiano, D.,  
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,  
Baldwin, D., Bandaruaike, D., Barber, M., Barnstead, M., Benahmed, P.,  
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
Cardenas, V., Carter, K., Cavazos, I., Ceasari, H., Centes, A.,  
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
Egan, A., Escoto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,  
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
Greggorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,  
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,  
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,  
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,  
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
Kowals, C., Kratt, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
Lorenshewari, L., Louisedge, H., Lozado, R.J., Lu, X., Ma, J.,  
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,  
Mangum, B., Mapa, P., Martin, K., Martin, R., Meenen, E.,  
Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,  
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
Morgan, N., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,  
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,



Nwaokelemeh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankoch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Resier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.O., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajda, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steidle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umanai, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Woodson, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission  
Unpublished  
2 (bases 1 to 224789)  
Worley, K.C.

Direct Submission  
Submitted (19-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 224789)

Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (13-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Nov 13, 2002 this sequence version replaced gi:22857292.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: GMSK  
Center clone name: CH230-237P17  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 209106 bases at least Q40  
Consensus quality: 213906 bases at least Q30  
Consensus quality: 216632 bases at least Q20  
Estimated insert size: 218750; sum-of-contigs estimation  
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 1 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 224789: contig of 224789 bp in length.

FEATURES  
source  
Location/Qualifiers  
1. 224789  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-237P17"

ORIGIN  
Query Match 47.2%; Score 200; DB 2; Length 224789;  
Best Local Similarity 80.9%; Pred. No. 1.3e-52;  
Matches 233; Conservative 0; Mismatches 55; Indels 0; Gaps 0;  
QY 137 CAGTCGTCGGGAGAGTGTCACAGGAGACTGTGTGTCGGAGAGTCCAGAGTCGGGCGAG 196  
DB 43369 CAGGTCCTCGGGAAGGTGTGTCAGGTGACTGCGGACCCAGGCGGAAACCCAGAGTCGGGCTG 43428  
QY 197 TGTGCTGCTTTTCATGTTGACGGGTGGACAACTCACCTGTCTAACTGTGTTGGTGAGAGCAACA 256  
DB 43429 TGTGCTGCTTTTCATGTTGACGGGTGGACAACTCACCTGTCTAACTGTGTTGGTGAGAGCAACA 256  
QY 257 GGCTCCAAAGGAAAGATTGTTTCGAGTTTGTGACTGGCACAGTACCTCTTTTCAGT 316  
DB 43489 GACCTCCGAAGGAAGAGGCTGTTTCGCGGTGTGTGACTGGCACAGTACCTCTTTTCAGT 43548  
QY 317 GGGAGGTTTCTGACTGTCGACCACTGTGTGCTTCTCTACGCTCGGCTGAGTGAAGTCAAGC 376  
DB 43549 GGGAGGTTTCTGACTGTCGACCACTGTGTGCTTCTCTACGCTCGGCTGAGTGAAGTCAAGC 376  
QY 377 CTGCGACTGCAGAGTGTGTGACCGCTCAGCATGAGTGCAGCACCAGCA 424  
DB 43609 CTGCGGCTGTGGAGTGTGTGACTGCGCAGCAGCGTCTACAGCACCAGCA 43656

RESULT 6  
AC123741/c  
LOCUS  
DEFINITION Mus musculus clone RP24-113A5, WORKING DRAFT SEQUENCE, 5 unordered pieces.  
ACCESSION AC123741  
VERSION AC123741.4 GI:30017756  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Mammalia; Eutheraia; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1 (bases 1 to 259236)  
Birren, B., Nusbaum, C. and Lander, E.  
Mus musculus, clone RP24-113A5  
Unpublished  
2 (bases 1 to 259236)  
Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menes, L., Mihova, T., Mieng, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliou, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
Direct Submission

TITLE



\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 43 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```

1 1199: contig of 1199 bp in length
1200: gap of unknown length
1300: contig of 1261 bp in length
2561: gap of unknown length
2660: contig of 1396 bp in length
2661: gap of unknown length
4057: contig of 1229 bp in length
4157: gap of unknown length
5385: contig of 1229 bp in length
5485: gap of unknown length
7183: contig of 1698 bp in length
7283: gap of unknown length
7284: contig of 2238 bp in length
9521: gap of unknown length
9522: contig of 1773 bp in length
11394: gap of unknown length
11395: contig of 1702 bp in length
11495: contig of 1702 bp in length
13197: gap of unknown length
13297: contig of 1528 bp in length
14824: gap of unknown length
14825: contig of 1750 bp in length
14925: gap of unknown length
16674: gap of unknown length
16675: contig of 1883 bp in length
16775: gap of unknown length
18657: contig of 1528 bp in length
18658: gap of unknown length
20286: contig of 2234 bp in length
20385: gap of unknown length
20386: contig of 2436 bp in length
22620: gap of unknown length
22621: contig of 2436 bp in length
25155: gap of unknown length
25256: gap of unknown length
27699: contig of 2444 bp in length
27700: gap of unknown length
29031: contig of 1232 bp in length
29032: gap of unknown length
31136: contig of 2005 bp in length
31236: gap of unknown length
33519: contig of 2283 bp in length
33619: gap of unknown length
35673: contig of 2054 bp in length
35674: gap of unknown length
37446: contig of 1673 bp in length
37447: gap of unknown length
37547: contig of 3213 bp in length
40760: gap of unknown length
40859: contig of 2419 bp in length
43278: gap of unknown length
43378: contig of 3178 bp in length
46556: gap of unknown length
46557: contig of 3829 bp in length
50485: gap of unknown length
50585: contig of 3074 bp in length
53659: gap of unknown length
53759: gap of unknown length
57160: contig of 3401 bp in length
57260: gap of unknown length
60085: contig of 2805 bp in length
60165: gap of unknown length
64163: contig of 3998 bp in length
64263: gap of unknown length
67791: contig of 3528 bp in length
67891: gap of unknown length
67992: contig of 4331 bp in length
72222: gap of unknown length
72223: contig of 4542 bp in length
76864: gap of unknown length
76865: contig of 4943 bp in length
81907: gap of unknown length
82007: contig of 5111 bp in length
82008:

```

```

* 87119 87218: gap of unknown length
* 87219 92188: contig of 4970 bp in length
* 92189 92888: gap of unknown length
* 92889 97635: contig of 5347 bp in length
* 97636 97735: gap of unknown length
* 97736 104390: contig of 6655 bp in length
* 104391 104490: gap of unknown length
* 104991 110209: contig of 5719 bp in length
* 110210 110309: gap of unknown length
* 110310 117135: contig of 6826 bp in length
* 117136 117235: gap of unknown length
* 117236 123407: contig of 6172 bp in length
* 123408 123507: gap of unknown length
* 123508 130670: contig of 7163 bp in length
* 130671 130770: gap of unknown length
* 130771 132225: contig of 8455 bp in length
* 132226 139325: gap of unknown length
* 139326 150961: contig of 11636 bp in length
* 150962 151061: gap of unknown length
* 151062 164366: contig of 13305 bp in length.

FEATURES
    Location/Qualifiers
        1..164366
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="2"
            /clone="RF11-788B17"
        1..1199
            /note="assembly_name:Contig23"
        1300..2560
            /note="assembly_name:Contig26"
        2661..4056
            /note="assembly_name:Contig32"
        4157..5385
            /note="assembly_name:Contig33"
        5486..7183
            /note="assembly_name:Contig35"
        7284..9521
            /note="assembly_name:Contig36"
        9622..11394
            /note="assembly_name:Contig37"
        11495..13196
            /note="assembly_name:Contig38"
        13297..14824
            /note="assembly_name:Contig39"
        14925..16674
            /note="assembly_name:Contig40"
        16775..18657
            /note="assembly_name:Contig41"
        18758..20285
            /note="assembly_name:Contig42"
        20386..22619
            /note="assembly_name:Contig43"
        22720..25155
            /note="assembly_name:Contig44"
        25256..27699
            /note="assembly_name:Contig45"
        27800..29031
            /note="assembly_name:Contig46"
        clone_end:77
        vector_side:right
        29132..31136
            /note="assembly_name:Contig47"
        31237..33519
            /note="assembly_name:Contig48"
        33620..35673
            /note="assembly_name:Contig49"
        35774..37446
            /note="assembly_name:Contig50"
        37547..40759
            /note="assembly_name:Contig51"
        40860..43278
            /note="assembly_name:Contig52"

```

```
misc_feature 43379..46556
              /note="assembly_name:Contig53"
misc_feature 46657..50485
              /note="assembly_name:Contig54"
misc_feature 50586..53659
              /note="assembly_name:Contig55"
misc_feature 53760..57160
              /note="assembly_name:Contig56"
misc_feature 57261..60065
              /note="assembly_name:Contig57"
misc_feature 60166..64163
              /note="assembly_name:Contig58"

Query Match      33.3%; Score 141; DB 2; Length 164366;
Best Local Similarity 100.0%; Pred.No. 1.2e-33;
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTTTCCAAAGAGCAACTAACAGTCACTTCTGCTGGGTATGGAGGAGCATGAGGAAGCTC 60
Db 77020 ATGTTTCCAAAGAGCAACTAACAGTCACTTCTGCTGGGTATGGAGGAGCATGAGGAAGCTC 77079

QY 61 TTTCATTGCTTTCTCTCTCTGCTGCTCCCATGCACTCATTTTGGAGGCAAAAGGATAAT 120
Db 77080 TTTCATTGCTTTCTCTCTGCTGCTCCCATGCACTCATTTTGGAGGCAAAAGGATAAT 77139

QY 121 CAGTTCATCTGGAACACAGGT 141
Db 77140 CAGTTCATCTGGAACACAGGT 77160

RESULT 8
AC010873/c
LOCUS          180602 bp      DNA      linear      PRI 07-NOV-2001
DEFINITION    Homo sapiens BAC clone RP11-112N16 from 2, complete sequence.
ACCESSION     AC010873
VERSION       AC010873.12 GI:13932726
KEYWORDS      HTG.
SOURCE        Homo sapiens (human)
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 180602)
AUTHORS       Sulston,J.E. and Waterston,R.
TITLE         Toward a complete human genome sequence
JOURNAL       Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE       99063792
PUBMED        9847074
REFERENCE     2 (bases 1 to 180602)
AUTHORS       Kyung,K. and Haakenson,W.
TITLE         The sequence of Homo sapiens BAC clone RP11-112N16
JOURNAL       Unpublished
REFERENCE     3 (bases 1 to 180602)
AUTHORS       Waterston,R.H.
TITLE         Direct Submission
JOURNAL       Submitted (25-SEP-1999) Genome Sequencing Center, Washington
              University School of Medicine, 4444 Forest Park Parkway, St. Louis,
              MO 63108, USA
REFERENCE     4 (bases 1 to 180602)
AUTHORS       Waterston,R.H.
TITLE         Direct Submission
JOURNAL       Submitted (08-MAY-2001) Genome Sequencing Center, Washington
              University School of Medicine, 4444 Forest Park Parkway, St. Louis,
              MO 63108, USA
REFERENCE     5 (bases 1 to 180602)
AUTHORS       Waterston,R.
TITLE         Direct Submission
JOURNAL       Submitted (09-AUG-2001) Department of Genetics, Washington
              University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE     6 (bases 1 to 180602)
AUTHORS       Waterston,R.
TITLE         Direct Submission
JOURNAL       Submitted (07-NOV-2001) Department of Genetics, Washington
              University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
```

COMMENT

On May 8, 2001 this sequence version replaced gi:13518243.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)  
----- Summary Statistics  
Center project name: H\_NH0112N16  
-----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is RP11-442L5, 200 bp overlap; the clone sequenced to the left is RP11-13K3, 200 bp overlap. Actual start of this clone is at base position 195 of RP11-112N16; actual end is at base position 180408 of RP11-112N16.

Polymorphic base differences occur in the overlap between RP11-112N16 and RP11-13K3.

The sequence from base position 151571 to 151695 and from position 151327 to 151337 is represented by a single strand as well as single chemistry.

FEATURES

Location/Qualifiers  
SOURCE  
1..180602  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/map="2"  
/clone="RP11-112N16"  
/clone\_lib="RPCI-11"  
repeat\_region 1..73  
/rpt\_family="MIR"  
repeat\_region 255..500  
/rpt\_family="L2"  
repeat\_region 501..921  
/rpt\_family="MaLR"  
repeat\_region 922..1317  
/rpt\_family="L2"  
repeat\_region 1331..2151  
/rpt\_family="L1"  
repeat\_region 2979..3101  
/rpt\_family="MIR"  
repeat\_region 3540..3870

```
/rpt_family="MaLR"
repeat_region 3871..5452
/rpt_family="MaLR"
repeat_region 5482..5901
/rpt_family="MaLR"
misc_feature 613..6703
/note="similar to EST A1141965 (NID:g3649422) ool7e06.xl"
613..6701
/note="similar to EST AW9623984 (NID:g8152820)"
613..6628
/note="similar to EST A1611205 (NID:g4620372) tw68h07.xl"
617..6701
/note="similar to EST A1830622 (NID:g5451293) wj10a04.xl"
617..6895
/note="similar to EST AA588635 (NID:g2402366) nj06d03.sl"
617..6674
/note="similar to EST A1090108 (NID:g3429167) qb28a11.xl"
618..6703
/note="similar to EST AW517645 (NID:g7155727) xt72c03.xl"
618..6701
/note="similar to EST AA358722 (NID:g2011041)"
618..6703
/note="similar to EST AA368086 (NID:g2020445)"
618..6703
/note="similar to EST AA902480 (NID:g3037603) ok70h12.sl"
618..6701
/note="similar to EST BF195923 (NID:g11083315)"
618..6703
/note="similar to EST AA358723 (NID:g2011042)"
618..6704
/note="similar to EST BE539677 (NID:g9768322)"
618..6703
/note="similar to EST AW383478 (NID:g6888241)"
618..6704
/note="similar to EST AW383489 (NID:g6888252)"
618..6761
/note="similar to EST BE739311 (NID:g10153389)"
618..6759
/note="similar to EST BF344750 (NID:g11292193)"
618..6759
/note="similar to EST BG163554 (NID:g12670257)"
618..6755
/note="similar to EST BE542005 (NID:g9770650)"
618..6761
/note="similar to EST BF086638 (NID:g10892257)"
618..6807
/rpt_family="(TG)n"
613..6704
/note="similar to EST BF086638 (NID:g10892257)"
613..6704
/note="similar to EST BF344750 (NID:g11292193)"
613..6704
/note="similar to EST BG163554 (NID:g12670257)"
613..6704
/note="similar to EST BE739311 (NID:g10153389)"
613..6705
/note="similar to EST BE542005 (NID:g9770650)"
613..6704
/note="similar to EST H58372 (NID:g1011204) yr20b10.rl"
613..6704
/note="similar to EST AA469946 (NID:g2197255) zt98g04.sl"
613..6749
/note="similar to EST AW382106 (NID:g6886765)"
613..6751
/note="similar to EST BE568877 (NID:g9812597)"
613..6736
/note="similar to EST AW896456 (NID:g8060661)"
613..6751
/note="similar to EST BG036141 (NID:g12430992)"
613..6742
/note="similar to EST T81123 (NID:g704008) yd24a05.rl"
613..6751
/note="similar to EST BE393709 (NID:g9339061)"
```

```
7346..7521
/note="similar to EST AW816566 (NID:g7909560)"
7349..7519
/note="similar to EST AW361153 (NID:g6865803)"
7353..7519
/note="similar to EST BG532258 (NID:g13523796)"
7420..7519
/note="similar to EST AA278542 (NID:g1919927) zs77h10.rl"
7429..7519
/note="similar to EST AW604121 (NID:g7308862)"
7440..7519
/note="similar to EST BF247499 (NID:g11623992)"
7545..8028

Query Match 33.3%; Score 141; DB 9; Length 180602;
Best Local Similarity 100.0%; Pred. No. 1.2e-33;
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGTTTCCAAAGGACCACTTAACAGTCACCTGCTGGTATGGAGGACGATGAGGAGCTC 60
Db 107961 ATGTTTCCAAAGGACCACTTAACAGTCACCTGCTGGTATGGAGGACGATGAGGAGCTC 107902
Qy 61 TTCTATTGCTTTCTCTCTGCTGCTGCCATGACCTCATTTGGAAGCAAAAAGGATAAT 120
Db 107901 TTCTATTGCTTTCTCTCTGCTGCTGCCATGACCTCATTTGGAAGCAAAAAGGATAAT 107842
Qy 121 CAGTTCATCTGGAACACGAGT 141
Db 107841 CAGTTCATCTGGAACACGAGT 107821

RESULT 9
AC096222/c
LOCUS AC096222 234528 bp DNA linear HTG 10-MAY-2003
DEFINITION Rattus norvegicus clone CH230-27C4, *** SEQUENCING IN PROGRESS ***.
ACCESSION AC096222
VERSION AC096222.6 GI:30522362
KEYWORDS HTG; HTGS PHASE2; HTGS DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 234528)
Muzny, D., Marie, E., Metzker, M., Lee, A., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Deigado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kwis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorensuhwa, L., Loulseg, H., Lozano, R.J., Lu, X., Ma, J.,
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, B., Montemayor, J., Moore, S.,
```

Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwankwelen, O., Okwundu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plommer, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Reilly, A., Reuter, M., Richards, S., Riggs, F., Rives, C., Rocky, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D., Sneed, A., Sodergren, E., Song, K.-Z., Sorelle, R., Sosa, J., Stearns, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission  
Unpublished  
2 (bases 1 to 234528)  
Worley, K.C.

Direct Submission  
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 234528)  
Rat Genome Sequencing Consortium.

Direct Submission  
Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On May 10, 2003 this sequence version replaced gi:24942280.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GERM  
Center clone name: CH230-27C4  
----- Summary Statistics  
Assembly program: Atlas 3.0;  
Consensus quality: 216647 bases at least Q40  
Consensus quality: 220163 bases at least Q30  
Consensus quality: 222500 bases at least Q20  
Estimated insert size: 232388; sum-of-contigs estimation  
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 1 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* been provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and

\* the accession number will be preserved.  
\* 1 234528: contig of 234528 bp in length.

FEATURES  
Location/Qualifiers  
1..234528  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-27C4"  
1..1930  
/note="wgs contig"  
misc\_feature  
229739..231173  
/note="wgs contig"  
misc\_feature  
232008..234528  
/note="wgs contig"

ORIGIN  
Query Match 26.1%; Score 110.6; DB 2; Length 234528;  
Best Local Similarity 86.5%; Pred. NO. 7.3e-24;  
Matches 122; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 1 ATGTTTCCAAAGACGACCTACAGTCACTTCTGCTGCTATGAGGAGCATGAGGAGCTC 60  
|||||  
DB 151980 ATGTTTCCAAAGAGTGATCTGCGCTCACTCACTGGGTTTCGAGGAGCATGAGGAGTCT 151921  
|||||  
QY 61 TTCTATTGCTTTCTCTCTTCTGCTGCTCCATGAGTCTATTTGGAAGCAAAAGGATAAT 120  
|||||  
DB 151920 TTCTAGTGTCTTCTCTCTGCTGCTCCCAAGCAGTCTATTTGGAAGGAGGAGGACAAAT 151861  
|||||  
QY 121 CAGTTCACTCGAAACACAGGT 141  
|||||  
DB 151860 CAGTTCACTCGAAACACAGGT 151840  
|||||

RESULT 10  
AC124806/c  
LOCUS  
AC124806  
DEFINITION  
Mus musculus chromosome 1 clone RP24-282F20 map 1, \*\*\* SEQUENCING  
IN PROGRESS \*\*\*, 6 unordered pieces.  
AC124806  
VERSION  
AC124806.4 GI:39930801  
KEYWORDS  
HTG: HTGS\_PHASE1; HTGS\_FULLTOP; HTGS\_ACTIVEFIN.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1 (bases 1 to 175131)  
AUTHORS  
Birren, B., Nusbaum, C. and Lander, E.  
TITLE  
Mus musculus chromosome 1, clone RP24-282F20  
JOURNAL  
Unpublished  
REFERENCE  
2 (bases 1 to 175131)  
AUTHORS  
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzgerald, M., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Huime, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Lakocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Mihov, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thoman, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
Direct Submission



JOURNAL Submitted (17-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 3 (bases 1 to 175131)

AUTHORS Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, M., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukagalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., DeArelano, K., Diaz, J.S., Dodge, K., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kellis, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Melidrim, J., Meneus, L., Mihova, T., Mlewa, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE Direct Submission

JOURNAL

COMMENT Submitted (16-DEC-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Dec 16, 2003 this sequence version replaced gi:29294265.

All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L26619

Center clone name: 282\_F\_20

-----

\* NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 4834: contig of 4834 bp in length  
 4835 4934: gap of 100 bp  
 4935 63418: contig of 58484 bp in length  
 63419 63518: gap of 100 bp  
 63519 67337: contig of 3819 bp in length  
 67338 67437: gap of 100 bp  
 67438 147955: contig of 80518 bp in length  
 147956 148055: gap of 100 bp  
 148056 168460: contig of 20405 bp in length  
 168461 168560: gap of 100 bp  
 168561 175131: contig of 6571 bp in length.

FEATURES  
 source  
 1..175131  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="1"  
 /map="1"  
 /clone="RP24-282F20"  
 /clone\_lib="RPCI-24 Male Mouse BAC"

## ORIGIN

Query Match 24.6%; Score 104.2; DB 2; Length 175131;  
 Best Local Similarity 83.7%; Pred. No. 8.1e-22;  
 Matches 119; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 1 ATGTTTCCAAAGACCAACCTAAACAGTCACTTGTGGTATGGAGGAGCATGAGGAAGCTC 60  
 |||||  
 Db 76673 ATGTTTCTGAGGAGTGATCTGGCTGTCACTCACTGGGTTCCAGGAGCATGAGGAAGCTC 76614  
 |||||  
 QY 61 TTTCTATTGCTTTCTCTCTCTGCTGTGCTCCATCGAGCTCATTTGGAAGGCAAAAGAGTAAT 120  
 |||||  
 Db 76613 TTTCTAGTGTCTTTCTCTCTCTGCTGTGCTCCCAAGCAGCTCATTTGGAAGGCAAGGACAAT 76554  
 |||||  
 QY 121 CAGTTTCATCTGGAACACAGGT 141  
 |||||  
 Db 76553 CAGTTTCCTCTGGAACACCGGT 76533  
 |||||

RESULT 11  
 AX747433 3053 bp mRNA linear PAT 20-JUN-2003  
 LOCUS AX747433 Sequence 958 from Patent EP1308459.  
 DEFINITION AX747433  
 ACCESSION AX747433  
 VERSION AX747433.1 GI:32131821  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1  
 AUTHORS Isogai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S., Yamamoto, J.I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R., Tanechika, I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahari, K. and Masuho, Y.

TITLE Full-length cDNA sequences  
 JOURNAL Patent: EP 1308459-A 958 07-MAY-2003;  
 Helix Research Institute (JP); Research Association for Biotechnology (JP)

FEATURES  
 source  
 1..3053  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"

ORIGIN

Query Match 18.5%; Score 78.6; DB 6; Length 3053;  
 Best Local Similarity 58.4%; Pred. No. 9.8e-14;  
 Matches 157; Conservative 0; Mismatches 109; Indels 3; Gaps 1;

QY 90 TGCAGCTCATTTGGAAGGCAAAAGGATAATCAGTTTCATCTGGAACCCAGGTCGTCGGGG 149  
 |||||  
 Db 393 TCCGCGCAGGCGGAGCGGCGGCCCCACCTCTATCTGTGGAAGACTGGTCCATGGGG 452  
 |||||

QY 150 AAGGTGTACAGGA---GACTGTGTGTCGCCGAGGAGTCCAGAGTGGCGCAGTGTGTTT 206  
 |||||  
 Db 453 CCGATGTATGGAGATGAATGTGTCGCCGAGGATCCAAACGAGGCTGTGTGTGTGC 512  
 |||||

QY 207 TCATGTTGACGGTGGACAAGTCACTGTCTACTGTGTGAGAGCAACAGGCTCCAAA 266  
 |||||  
 Db 513 TCATGTGGAGGATGGACTACACTGCATACATTAACCTGTAAGCAGCCGAGAGACCAATAA 572  
 |||||

QY 267 GGAAGAAGTGTGTTCCGAGTTTGTGACTGGCAGTGCACCTCTTTTCAGTGGGAGGTTTC 326  
 |||||  
 Db 573 CCAGCAGATTGTTTCAAGTTTCCGATTGGCACAAGAGTTGTACACTGGAGACTGGG 632  
 |||||

QY 327 TGATGGCACCACTGTGTGTGTTTCCTT 355  
 |||||  
 Db 633 ACCTTGAATCGTGTGAGCCCGTGATT 661  
 |||||

RESULT 12  
 AK092252 3053 bp mRNA linear PRI 15-JUL-2002  
 LOCUS AK092252 Homo sapiens cDNA FLJ34933 fis, clone NT2RF005669.  
 DEFINITION AK092252  
 ACCESSION AK092252  
 VERSION AK092252.1 GI:21750796  
 KEYWORDS oligo capping; fis (full insert sequence).  
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1  
AUTHORS Iehibashi, T., Kanehori, K., Yosida, M., Watanabe, S., Ishida, S.,  
Ono, Y., Hotuta, T., Hiraoka, S., Murakawa, K., Takiguchi, S.,  
Kusano, J., Watanabe, M., Fujimori, K., Tanai, H., Ishida, M.,  
Yamashita, H., Chiba, Y., Sugiyama, T., Irie, R., Otsuki, T., Sato, H.,  
Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y.,  
Saito, K., Nishikawa, T., Kimura, K., Matsuo, K., Nakamura, Y.,  
Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Takehashi-Fujii, A.,  
Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S.,  
Nagatani, K., Masuho, Y., Nagai, K. and Isogai, T.  
NEDO human cDNA sequencing project

UNPUBLISHED  
2 (bases 1 to 3053)  
Isogai, T. and Yamamoto, J.  
Direct Submission  
Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7  
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan  
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)  
NEDO human cDNA sequencing project supported by Ministry of  
Economy, Trade and Industry of Japan; cDNA full insert sequencing:  
Research Association for Biotechnology (RAB); cDNA library  
construction: Helix Research Institute (HRI) (supported by Japan  
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,  
HRI, and Biotechnology Center, National Institute of Technology and  
Evaluation; clone selection for full insert sequencing: HRI and  
RAB; annotation: HRI and RAB.  
Location/Qualifiers  
1...3053  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="NT2RP7005669"  
/cell\_line="NT2"  
/cell\_type="teratocarcinoma"  
/clone\_lib="NT2RP7"  
/notes="cloning vector: pME18SPL3-mRNA from NT2 neuronal  
precursor cells after 5-weeks retinoic acid (RA)  
induction."

ORIGIN  
Query Match 18.5%; Score 78.6; DB 9; Length 3053;  
Best Local Similarity 58.4%; Pred. No. 9.8e-14;  
Matches 157; Conservative 0; Mismatches 109; Indels 3; Gaps 1;  
QY 90 TGCAGCTCATTTGGAAGCGCAAAAGGATATACGTTTCATCTCGAACCAGGTCCTGGGG 149  
Db 393 TCGCGCGAGGCGGAGCGGCGGCGGCCCCCTCTATCTGTGAAGACTGTCCTCAATGGG 452  
QY 150 AAGGTGTACAGGA---GACTGTGTCTCCGAGGAGTCCAGAGTCGGGCGAGTGTGGTGT 206  
Db 453 CCGATGTATGGGAGATGAATGTGTCTCCGAGGCGCATCAAACGAGGCTGTGGTGTGC 512  
QY 207 TCATGTTGACGGGTGCACAGTCACTGTCTAATCTGTGTGTGAGAGCAACAGGCTCCAA 266  
Db 513 TCATGTGAGGGATGGACTACATGCACTATCACTGTAACCTGAACGCGCCGAGAGACCAATA 572  
QY 267 GGAAGAAGTGTGTTTCCAGTTTGTGACTGGCAGAGTGACTCTTTTCAGTGGGAGGTTTC 326  
Db 573 CCACGAGATTGTTTCAAGTTTGGATTGGCACAAGAGTTGTACGACTGGAGACTGGG 632  
QY 327 TGACTGGCACCAGTGTGTGTGTTGTTCC 355  
Db 633 ACCTTGAATCGGTGTACAGCCCGTGATT 661

RESULT 13  
AR338841  
LOCUS AR338841 6378 bp DNA linear PAT 17-AUG-2003  
DEFINITION Sequence 332 from patent US 6569662.  
ACCESSION AR338841

AR338841.1 GI:33725698  
KEYWORDS Unknown.  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 6378)  
AUTHORS Tang, Y.T., Zhou, P. and Drmanac, R.T.  
TITLE Nucleic acids and polypeptides  
JOURNAL Patent: US 6569662-A 332 27-MAY-2003;  
FEATURES Location/Qualifiers  
1..6378  
source /organism="unknown"  
/mol\_type="genomic DNA"  
ORIGIN  
Query Match 17.9%; Score 76; DB 6; Length 6378;  
Best Local Similarity 59.8%; Pred. No. 7.2e-13;  
Matches 146; Conservative 0; Mismatches 95; Indels 3; Gaps 1;  
QY 115 GATATACAGTTTCATCTGGAAACCAAGTCCGTCGGGGAAGGTGTACAGGA---GACTGTGGT 171  
Db 229 GGTATTCGGAAGCTTCTGAGAAAGTCCATGGGCGCGATGTATGGGAGATGAATGTGGT 288  
QY 172 CCGGAGGAGTCCAGAGTCGGGCGAGTGTGGTGTTCATGTTCACGGGTGGACAGTCAC 231  
Db 289 CCGGAGGAGTCCAAACGAGGCGCTGTGGTGTCTCATGTGGAGGATGGACTACACTG 348  
QY 232 CTGTCTAATCTGTGTGTGAGCAACAGGCTCCAAAGGAAGAAGTTGTTCCGAGTTGT 291  
Db 349 CATACTAATCTGTAAGCAGCGCGAGAGACCCCAATAACAGCAGCAATTCGTTCAAAAGTTGC 408  
QY 292 GACTGGCACAGTGACCTCTTTTCAGTGGGAGGTTTCTGACTGGCACCAGTGTGCTGTT 351  
Db 409 GATTGGCACAAAGATGTTGACTGTGAGACTGGGACCTTGGAAATCAGTGTACGCCCGTG 468  
QY 352 CCGT 355  
Db 469 ATTT 472

RESULT 14  
AR338841  
LOCUS AR338841 6373 bp DNA linear PAT 22-FEB-2001  
DEFINITION Sequence 1 from Patent WO0105971.  
ACCESSION AR338841  
VERSION AR338841.1 GI:33159395  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Shimkets, R.A. and Fernandes, E.  
TITLE Secreted polypeptides and corresponding polynucleotides  
JOURNAL Patent: WO 0105971-A 1 25-JAN-2001;  
FEATURES Location/Qualifiers  
1..6373  
source /organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
235..5001  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC32422.1"  
/db\_xref="GI:13159396"  
/translation="MGDECGPGIOTBAVNCVHVEGHTTLHTNCKOARPNNOCNCFK  
VCDWHKELYDWRGLPWNQCPVSKSELEKICKEGQVREIACIQKDKIIPAD  
IICSYFEPKPLLEOACLIPOQDCIVSEFSAMSECSKTCGSGLOHRTHRHVAVPFGG  
SGCNLTFFQVCSGPCAEELRSLVHGWPMSCMPSHPSQVROARRRGNKREKOR  
SKGYDPEARELIKKNRNRRNQENKYWDIQIQTREVMCINTKTKAADSFCQQ  
EKLPMTFQSCVITKECVSEWSWSPCSKTDHNVSPAGTRVTRTRTIRPPIGSEKEC

```

PEPEKEPCLSGDGVPCATYGTWTTWTECRVDPDLLSQDQKRGNGTALCGGGIOT
REVIYCVANENLQSLSTHKNKASKPMDLKLTGTPINTTQOLCHIPCTECPYSPMS
AWGCTPENCNDQKGGKFKLRKITNEPTGGSGVTCNCPHLEAIPCEBPACYDVK
AVRLGCEPDNGCEKGTQVEVVCINSDEGEVDRQCDRAIPPIVACDAPCKOC
VLSTWSTSSCSHTSGTKTEGQILARASILAYAGEGGICRPNSSALQOVRSCNEHP
CTVYHWTGPMQSCITTSVSEFNTTWTNGEASCSVGMOTRKVLCVRVNVGVGPXK
CPESLRPBTVRPCLLPCKKDCIPIYPSDWTSCPSCKEGSSIRKQSRHVVIIOLPAN
GARDCTDIYEXKAEAFQACQSYRWKTHKRCOLVPSQSPQAQCEGPGRA
RAITCKQGGQAGHECLOVAGFPALTOACQPCDDCQLTWSKFPSSCNGCGAV
RTKRKTLVKSCKEKKXNHLPIETQYCDKYNAPVGNMSDCLIPGKVEVLL
GNKVGDIKECGQYTYQAMACYDQNGRLVETSRNCHGYIEEACIIPCPDCKLSW
SNWSSCSGSGVRSKMLREKPYNGRCPCKLDHVNQAOVYEVVPCSDCNQYSLW
VTEPWSICVTFVNRGCEGVOTRKVRCMONTADGSEHVEDYLCDPERPMPGSAV
CKLCPEDCVISEMGPWTQCVLPNCNOSFRORSADP IQPADEGRSCPNAVEKEPCNL
KNCHYDYNVTDHSTQJLSEKAVCGNIGIKTRMLDVCVRSKGLKCEALGLEKRM
QNTSCMEVPCVQLSWSPWSECSQTCGLTGKMRRTVTQFPQDGRPCPSLMDQ
SKPCVPKPCYQVQSCQVQVQACQEGTRTNSICVSDGSDADPFSKYVDEEFC
DIELIIDGNKNVLESCQPCQDYLKDWSSWLSCLTVCNGEDJGFGGIQVTRSP
VUIQELNHLQCEPMLETKSCYDQCYEYKMASANKGSRVTVCORSQDGINVTGC
LVMSOPDADRSNPPCSOPHSVCSKTCHEEGYEVMSNSNTLEOCTILIPVVLPT
MEDKGDVKTSPRAVHTQPSNPAGRGHTWLPQPGRLKTYWYGVGAAGAFVLLIF
IVSMYILACKPKPKPQRQNNRLKPLIAYDGDADM"

```

## .RIGIN

```

Query Match 17.8%; Score 75.4; DB 6; Length 6373;
Best Local Similarity 62.0%; Pred. No. 1.1e-12;
Matches 137; Conservative 0; Mismatches 81; Indels 3; Gaps 1;

>Y 138 AGTTCGTCGGGAAGGTGTACAGGA---GACTGTGTCGCGAGAGTCCAGATCGGGC 194
>b ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
>Y 216 AGTTCATCGGGCCGATGATGCGGAGATGAATGTGTCGCGAGGATCCAAAGCGGC 275
>b ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
>Y 195 AGTGTGTTTTCATGTTGACGCGGTGACAGTCACTGTCTACTGTGTTGAGAGCAA 254
>b ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
>Y 276 TGTGTGTTGTCATGTGAGGAGTGGACTACATGTGCTACTTACTTACGAGCGCGA 335
>b ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
>Y 255 GAGGCTCCAAAGGAAGTGTGTTTCGAGTTTGTGACTGGCACAGTGCCTTTTCA 314
>b ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
>Y 336 GAGACCCATACACGAGCAATGTGTTCAAGTTTGGATTGGCACAAAGAGTTGTACGA 395
>b ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
>Y 315 GTGGGAGGTTTCTGACTGSCACCACCTGTGCTGCTTCTT 355
>b ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
>Y 396 CTGGAGACTGGGACCTTGAATCAGTGTGACGCGCGTATT 436

```

```

RESULT 15
.AC004614/c
DEFINITION Homo sapiens PAC clone RP5-1168D11 from 7, complete sequence.
ACCESSION AC004614
VERSION AC004614.2 GI:21321876
SOURCE HTG.
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Sulston,J.B. and Wilson,R.
1 (bases 1 to 112456)
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
PUBMED 9847074
2 (bases 1 to 112456)
Blandford,M. and Lacy,M.
The sequence of Homo sapiens PAC clone RP5-1168D11
Unpublished (2001)
3 (bases 1 to 112456)
Waterston,R.
Direct Submission
Submitted (24-APR-1998) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
4 (bases 1 to 112456)
Waterston,R.

```

```

TITLE Direct Submission
JOURNAL Submitted (03-FEB-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 5 (bases 1 to 112456)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (04-JUN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 6 (bases 1 to 112456)
AUTHORS Wilston,R.
TITLE Direct Submission
JOURNAL Submitted (15-OCT-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Jun 4, 2002 this sequence version replaced gi:3080662.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: sapiens@wustl.edu
----- Summary Statistics
-----
Center project name: H_D1168D11

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

## MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GTS/CHR7>, send mailto:egreen@nhgri.nih.gov, or see <http://genome.wustl.edu>

## SOURCE INFORMATION:

This clone was derived from human PAC library RPCI-5, prepared by Pieter de Jong and coworkers at <http://www.chori.org> using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor. The clone may be obtained either from Genome Systems, Inc. (<http://www.genomesystems.com>) or Research Genetics, Inc. (<http://www.resgen.com>); or from Pieter de Jong.

```

VECTOR: PCVPAC2. Location/Qualifiers
1. 112456
   /organism="Homo sapiens"
   /mol_type="genomic DNA"
   /db_xref="taxon:9606"
   /chromosome="7"
   /map="7"
   /clone="RP5-1168D11"
   /clone_lib="RPCI-5"
   /rpt_family="ERVL"
   /rpt_family="MAlR"
   /rpt_family="MIR"
   /rpt_family="GA-rich"
   /rpt_family="5120"
   /rpt_family="6558"

```

```
repeat_region /rpt_family="MIR"
repeat_region 7807.7846
repeat_region /rpt_family=" (TCTA)n"
repeat_region 7892.8181
repeat_region /rpt_family="Alu"
repeat_region 9128.9156
repeat_region /rpt_family="AT_rich"
repeat_region 9263.9974
repeat_region /rpt_family="L1"
repeat_region 10035.10061
repeat_region /rpt_family=" (CA)n"
repeat_region 10062.10941
repeat_region /rpt_family="L1"
repeat_region 10934.11278
repeat_region /rpt_family="L1"
repeat_region 11279.12784
repeat_region /rpt_family="L1"
repeat_region 12786.13758
repeat_region /rpt_family="L1"
repeat_region 13764.14207
repeat_region /rpt_family="L1"
repeat_region 14208.14564
repeat_region /rpt_family="MaLR"
repeat_region 14565.16156
repeat_region /rpt_family="L1"
repeat_region 16362.16948
repeat_region /rpt_family="ERV1"
repeat_region 17800.17956
repeat_region /rpt_family="L1"
repeat_region 17960.18238
repeat_region /rpt_family="MaLR"
repeat_region 18239.22161
repeat_region /rpt_family="L1"
repeat_region 22591.23172
repeat_region /rpt_family="MaLR"
repeat_region 23327.23381
repeat_region /rpt_family=" (TA)n"
repeat_region 23449.23510
repeat_region /rpt_family="GA-rich"
repeat_region 23586.24083
repeat_region /rpt_family="L1"
repeat_region 24084.24650
repeat_region /rpt_family="ERV1"
repeat_region 24651.24722
repeat_region /rpt_family="L1"
repeat_region 24735.24951
repeat_region /rpt_family="MIR"
repeat_region 25267.25287
repeat_region /rpt_family="AT_rich"
repeat_region 25317.25418
repeat_region /rpt_family="MIR"
repeat_region 25419.25705
repeat_region /rpt_family="Alu"
repeat_region 25706.25762
repeat_region /rpt_family="MIR"
repeat_region 26548.26594
repeat_region /rpt_family="AT_rich"
repeat_region 27187.27491
repeat_region /rpt_family="Alu"
repeat_region 29221.29497
repeat_region /rpt_family="L2"
repeat_region 30155.30315
repeat_region /rpt_family="MER1_type"
repeat_region 30941.31320
repeat_region /rpt_family="MaLR"
repeat_region 31385.31563
repeat_region /rpt_family="MER1_type"
repeat_region 32215.32484
repeat_region /rpt_family="Alu"
repeat_region 32485.32833
repeat_region /rpt_family="L1"
repeat_region 32841.33720
repeat_region /rpt_family="L1"
```

```
repeat_region 33721.34024
/rpt_family="Alu"
repeat_region 34025.34467
/rpt_family="L1"
repeat_region 35276.35443
/rpt_family="MIR"
repeat_region 35446.35919
/rpt_family="L1"
repeat_region 35920.36041
/rpt_family="Alu"
repeat_region 36120.36345
/rpt_family="L1"
repeat_region 36346.36613
/rpt_family="Alu"
repeat_region 36614.36827
/rpt_family="L1"
repeat_region 37155.37475
/rpt_family="L1"
repeat_region 37657.37711
/rpt_family="L2"
repeat_region 37884.37963
/rpt_family="L2"
repeat_region 38009.38226
/rpt_family="L2"
repeat_region 38267.38565
/rpt_family="Alu"
misc_feature 38295.38562
/note="CpG_island ($CC=60.1, O/e=0.76, #CPGs=19)"

Query Match 17.8; Score 75.4; DB 9; Length 112456;
Best Local Similarity 62.0%; Pred. No. 1.4e-12;
Matches 137; Conservative 0; Mismatches 81; Indels 3; Gaps 1;

QY 138 AGGTCCGTGGGGAGGTCTACAGGA---GACTGTGGTCCCGGAGGAGTCCAGAGTCGGGC 194
Db 100793 AGTCCATGGGGCCGATGTATGGGAGATGAATGTGTCGGAGGCATCCAAACGAGGCG 100734

QY 195 AGTGTGGTGTCTTCATGTGTGACGGGTGGACAAAGTCACCTGTCTAACTGTGTGTGAGAGCAA 254
Db 100733 TGTGTGGTGTCTCATGTGTGAGGGGATGACTACACTGCATCTAACTGTAAAGCAGGCGA 100674

QY 255 CAGGCTCCAAAGGAAAGAGTGTTCGAGTTGTGACTGGCAGACAGTACCTTTCA 314
Db 100673 GAGACCCAATAACACAGAGAATTGTTCAAAGTTTGGGATTGGCACAAGAGTTGTACGA 100614

QY 315 GTGGGAGGTTCCTGACTGGCCACCACTGTGTCTTGTTCCTT 355
Db 100613 CTGGAGACTGGGACCTTGGATCAGTGTACGCCGTGATTT 100573
```

Search completed: February 23, 2004, 17:26:07  
Job time : 1598.77 secs

Result No.	Query			ID	Description
	Score	Match	Length		
1	285.8	67.6	5847	9	Ade07142 Novel
2	285	67.2	744	4	Aak343667 Human
3	285	67.2	744	4	Aak17814 Human
4	264.4	62.4	477	4	Aak30658 Human
5	264.4	62.4	477	4	Aak05095 Human
6	78.6	18.5	3053	9	Ade62804 Human
7	76	17.9	2326	3	AAC76705 Human
8	76	17.9	6378	4	AA158451 Human
9	76	17.9	6378	8	AdeB48422 Novel
10	75.4	17.8	1956	4	AA134372 Probe
11	75.4	17.8	1956	4	AB528046 Human
12	75.4	17.8	6373	4	AAC30188 Clon
13	75.4	17.8	6373	7	ACCT2162 Human
14	75.4	17.8	6373	7	ACCT2171 Human
15	75.4	17.8	6373	7	ACCT2168 Human
16	74.8	17.6	6317	6	AAC38694 Human
17	73.8	17.4	833	4	AA147548 Probe
18	73.8	17.4	833	4	AB541098 Human
19	73.8	17.4	921	7	ACCT2173 Human
20	73.8	17.4	4647	7	ACCT2167 Human
21	73.8	17.4	4679	7	ACCT2166 Human
22	73.8	17.4	4810	7	ACCT2174 Human
23	73.8	17.4	4891	7	ACCT2165 Human

CC proteins. The DNA and protein sequences of the invention are useful as:  
 CC markers for tissues in which the corresponding protein is preferentially  
 CC expressed; as molecular weight markers on gels; as chromosome markers or  
 CC tags; to identify chromosomes or to map related gene positions; and to  
 CC compare with endogenous DNA sequences in patients to identify potential  
 CC genetic disorders. The present DNA sequence represents a gene of the  
 CC invention.

XX  
 SQ Sequence 5847 BP; 1607 A; 1282 C; 1453 G; 1505 T; 0 U; 0 Other;  
 Query Match 67.6%; Score 286.8; DB 9; Length 5847;  
 Best Local Similarity 99.3%; Pred. No. 1.6e-81;  
 Matches 288; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 135 ACCAGGTCGCGGGAAGGTGTACAGGAGACTGTGTCGCCGAGGAGTCCAGAGTCGGGC 194  
 Db 33 ACATGTCGCGTGGGGAAGGTGTACAGGAGACTGTGTCGCCGAGGAGTCCAGAGTCGGGC 92  
 QY 195 AGTGTGGTGTTCATGTGTGACGGGTGGACAAGTCACCTGTCTAACTGTGTGTGAGAGCAA 254  
 Db 93 AGTGTGGTGTTCATGTGTGACGGGTGGACAAGTCACCTGTCTAACTGTGTGTGAGAGCAA 152  
 QY 255 CAGGCTCCAAAGGAAGAAAGTGTTCGCCGAGTTTGTGACTGGCACAGTGCACCTCTTTCA 314  
 Db 153 CAGGCTCCAAAGGAAGAAAGTGTTCGCCGAGTTTGTGACTGGCACAGTGCACCTCTTTCA 212  
 QY 315 GTGGAGGTTTCTGACTGGCACCACTGTGTCTTTCCTTACGCTCGCGTGAAGTCAA 374  
 Db 213 GTGGAGGTTTCTGACTGGCACCACTGTGTCTTTCCTTACGCTCGCGTGAAGTCAA 272  
 QY 375 GCCTCGGAAGTGTGTGACGGGTGACATGCACTGCACTGCACTGCACTGCACTGCACTGCA 424  
 Db 273 GCCTCGGAAGTGTGTGACGGGTGACATGCACTGCACTGCACTGCACTGCACTGCACTGCA 322

RESULT 2  
 AAK43667  
 ID AAK43667 standard; DNA; 744 BP.  
 XX  
 AC AAK43667;  
 XX  
 DT 06-NOV-2001 (first entry)  
 XX  
 DE Human bone marrow expressed single exon probe SEQ ID NO: 18224.  
 XX  
 KW Human; bone marrow expressed exon; gene expression analysis; probe;  
 KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FN WO200157276-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US000668.  
 XX  
 PR 04-FEB-2000; 2000US-0180312P.  
 XX  
 PR 26-MAY-2000; 2000US-0207456P.  
 XX  
 PR 30-JUN-2000; 2000US-00608408.  
 XX  
 PR 03-AUG-2000; 2000US-00632366.  
 XX  
 PR 21-SEP-2000; 2000US-0234687P.  
 XX  
 PR 27-SEP-2000; 2000US-0236359P.  
 XX  
 PR 04-OCT-2000; 2000GB-00024253.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-488900/53.  
 XX  
 PT Human genome-derived single exon nucleic acid probes useful for analyzing  
 PT gene expression in human bone marrow.  
 XX

PS Example 4; SEQ ID NO 18224; 658pp + Sequence Listing; English.  
 XX  
 CC The present invention provides a number of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC bone marrow. They can be used to measure gene expression in bone marrow  
 CC samples, which may enable the improved diagnosis and treatment of cancers  
 CC such as lymphoma, leukaemia and myeloma. The present sequence is one of  
 CC the probes of the invention

XX  
 SQ Sequence 744 BP; 181 A; 167 C; 201 G; 195 T; 0 U; 0 Other;  
 Query Match 67.2%; Score 285; DB 4; Length 744;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-81;  
 Matches 285; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 140 GTCCGTGGGGAAGGTGTACAGGAGACTGTGTCGCCGAGGAGTCCAGAGTCGGCAGTGT 199  
 Db 1 GTCCGTGGGGAAGGTGTACAGGAGACTGTGTCGCCGAGGAGTCCAGAGTCGGCAGTGT 60  
 QY 200 GGTGTTTTCATGTGTGACGGGTGGACAAGTCACTGTCTAACTGTGTGTGAGAGCAACAGGC 259  
 Db 61 GGTGTTTTCATGTGTGACGGGTGGACAAGTCACTGTCTAACTGTGTGTGAGAGCAACAGGC 120  
 QY 260 CTCCTAAAGCAAGAAAGTGTTCGCCGAGTTTGTGACTGGCACAGTGCACCTCTTTCA 319  
 Db 121 CTCCTAAAGCAAGAAAGTGTTCGCCGAGTTTGTGACTGGCACAGTGCACCTCTTTCA 180  
 QY 320 AGGTTTTCATGTGTGACGGCACCACTGTGTCTTTCCTTACGCTCGCGTGAAGTCAAGCTC 379  
 Db 181 AGGTTTTCATGTGTGACGGCACCACTGTGTCTTTCCTTACGCTCGCGTGAAGTCAAGCTC 240  
 QY 380 GCACTGCAAGTGTGTGACGGGTGACATGCACTGCACTGCACTGCACTGCACTGCACTGCA 424  
 Db 241 GCACTGCAAGTGTGTGACGGGTGACATGCACTGCACTGCACTGCACTGCACTGCACTGCA 285

RESULT 3  
 AAK17814  
 ID AAK17814 standard; DNA; 744 BP.  
 XX  
 AC AAK17814;  
 XX  
 DT 05-NOV-2001 (first entry)  
 XX  
 DE Human brain expressed single exon probe SEQ ID NO: 17805.  
 XX  
 KW Human; brain expressed exon; gene expression analysis; probe; microarray;  
 KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer;  
 XX ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FN WO200157275-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US000667.  
 XX  
 PR 04-FEB-2000; 2000US-0180312P.  
 XX  
 PR 26-MAY-2000; 2000US-0207456P.  
 XX  
 PR 30-JUN-2000; 2000US-00608408.  
 XX  
 PR 03-AUG-2000; 2000US-00632366.  
 XX  
 PR 21-SEP-2000; 2000US-0234687P.  
 XX  
 PR 27-SEP-2000; 2000US-0236359P.  
 XX  
 PR 04-OCT-2000; 2000GB-00024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-483446/52.  
 XX  
 PT Single exon nucleic acid probes for analyzing gene expression in human



```

XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human bone marrow.
XX
XX Example 4; SEQ ID NO 5215; 558pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is one of the probes of the
XX invention
XX
XX Sequence 744 BP; 181 A; 167 C; 201 G; 195 T; 0 U; 0 Other;
XX
XX Query Match 67.2%; Score 285; DB 4; Length 744;
XX Best Local Similarity 100.0%; Pred. No. 2.5e-81;
XX Matches 285; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 140 GTCCGTTGGGAGGTTGACAGAGACTGTGTCGCGAGGAGTCCAGAGTCCGCGAGTGT 199
XX 1 GTCCGTTGGGAGGTTGACAGAGACTGTGTCGCGAGGAGTCCAGAGTCCGCGAGTGT 60
XX
XX 200 GGTGTTTTCATGTTGACGGGTGGACAAAGTCACCTGTCTAACTGTGTTGAGAGCAACAGGC 259
XX 61 GGTGTTTTCATGTTGACGGGTGGACAAAGTCACCTGTCTAACTGTGTTGAGAGCAACAGGC 120
XX
XX 260 CTCCTCAAGGAAAGAGTTGTTTCCGAGTTGTGACTGGCAGTGCACCTCTTTCAAGTGGG 319
XX 121 CTCCTCAAGGAAAGAGTTGTTTCCGAGTTGTGACTGGCAGTGCACCTCTTTCAAGTGGG 180
XX
XX 320 AGTGTTCCTGACTGGCACCACCTGTGCTGTTGTTTCCCTTACGCTCGCGGTGAAGTCAGCCTC 379
XX 181 AGTGTTCCTGACTGGCACCACCTGTGCTGTTGTTTCCCTTACGCTCGCGGTGAAGTCAGCCTC 240
XX
XX 380 GGACTGCAGAGTGTGAGCGGCTCAGCATGAGTGGACTGCAGCACCGGA 424
XX 241 GGACTGCAGAGTGTGAGCGGCTCAGCATGAGTGGACTGCAGCACCGGA 285
XX
XX RESULT 4
XX AK30658
XX D AKK05095 standard; DNA; 477 BP.
XX
XX AKK05095;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human bone marrow expressed single exon probe SEQ ID NO: 5215.
XX
XX Human; bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
XX Homo sapiens.
XX
XX WO200157276-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000669.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488900/53.
XX
XX
XX Query Match 62.4%; Score 264.4; DB 4; Length 477;
XX Best Local Similarity 85.3%; Pred. No. 9e-75;
XX Matches 295; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
XX
XX 53 GGAAGCTCTTTCTATTGCTTTCTCTCTTGTGTCCTCCATGCGAGCTCATTTGGAAAGGCAAAA 112
XX 132 GGAAGCTCTTTCTATTGCTTTCTCTCTTGTGTCCTCCATGCGAGCTCATTTGGAAAGGCAAAA 191
XX
XX 113 AGGATAATCATGTTTCATCTGGAACACAGGTCCTGCGGAGAGTCTACAGAGACTGTGGTC 172
XX 192 TTAACATCTCTGTTTCTGCTGTTAGGTCCTGCGGAGAGTGTACAGAGACTGTGGTC 251
XX
XX 173 CCGGAGAGGTCAGAGTCGCGGAGTGTGTTGTTTTCATGTTGACGGTGGACAAAGTCACC 232
XX 252 CCGGAGAGGTCAGAGTCGCGGAGTGTGTTGTTTTCATGTTGACGGTGGACAAAGTCACC 311
XX
XX 233 TGCTTAACCTGCTGTGAGAGCAACAGAGCTCCCAAGAGAGAGTCTTCCGAGTTTGTG 292
XX 312 TGCTTAACCTGCTGTGAGAGCAACAGAGCTCCCAAGAGAGAGTCTTCCGAGTTTGTG 371
XX
XX 293 ACTGGCACAGTGCACCTCTTTTTCAGTGGGAGGTTTCTGACTGGCACCACCTGTGCTGTTC 352
XX 372 ACTGGCACAGTGCACCTCTTTTTCAGTGGGAGGTTTCTGACTGGCACCACCTGTGCTGTTC 431
XX
XX 353 CTTACGCTCGCGTGAAGTCAAGCTCGGACTCGGACTCGAGTGTGTGAC 398
XX 432 CTTACGCTCGCGTGAAGTCAAGCTCGGACTCGGACTCGAGTGTGTGAC 477
XX
XX RESULT 5
XX AAK05095
XX ID AAK05095 standard; DNA; 477 BP.
XX
XX AC AAK05095;
XX
XX DT 05-NOV-2001 (first entry)
XX
XX DE Human brain expressed single exon probe SEQ ID NO: 5086.
XX
XX KW Human; brain expressed exon; gene expression analysis; probe; microarray;
XX KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer;
XX KW ss.
XX
XX OS Homo sapiens.
XX
XX FN WO200157275-A2.
XX
XX PD 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000667.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488900/53.
XX
```





XX Novel nucleic acids and polypeptides, useful for treating disorders such  
 PT as central nervous system injuries.  
 XX  
 XX  
 PS Claim 1; SEQ ID NO 654; 10078pp; English.  
 XX  
 CC The invention relates to human nucleic acids (AA157798-AA161369) and the  
 CC encoded polypeptides (AA158642-AA162213) with nootropic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
 CC C.N.S disorders. Note: The sequence data for this patent did not form  
 CC part of the printed specification  
 XX  
 SQ Sequence 6378 BP; 1823 A; 1420 C; 1619 G; 1515 T; 0 U; 1 Other;

Query Match 17.9%; Score 76; DB 4; Length 6378;  
 Best Local Similarity 59.8%; Pred. No. 1.7e-13;  
 Matches 146; Conservative 0; Mismatches 95; Indels 3; Gaps 1;  
 QY 115 GATATCATGTTTCATCTGGAACACAGTCCGTGGGAGGTGTACAGGA---GACTGTGGT 171  
 Db 229 GGTAAATGGGAAAGCTTCTGAGAAAGTCCATGGGCCGATGTATGGGAGATGAATGTGT 288  
 QY 172 CCCGAGGAGTCCAGAGTCGGGCAGTGTGTTTTCATGTTGACGGGTGGACAAGTCA 231  
 Db 289 CCCGAGGATCCAAACGAGGGCTGTGTGTGTCTCATGTGGAGGATGACTACACTG 348  
 QY 232 CTGCTTAAGTCTGGTGAGCAACAGGCCTCCAAAGGAAGATTGTTCCGAGTTTGT 291  
 Db 349 CATACTAAGTCTAAGCAGCCGCGAGAGACCCCAATAACCAAGCAGAGATTGTTTCAAAAGTTGC 408  
 QY 292 GACTGGCAGTGTGAGTCTTTTCAGTGGGAGGTTTCTGACTGGCAGTGTGTTGTT 351  
 Db 409 GATTGGCAAAAGAGTGTACACTGTGAGACTGGGAGCTTGGAAATCAGTGTACGCCGTG 468  
 QY 352 CCTT 355  
 Db 469 ATTT 472

RESULT 9  
 ADB48422  
 ID ADB48422 standard; cDNA; 6378 BP.

AC ADB48422;  
 XX  
 DT 04-DEC-2003 (first entry)  
 XX  
 DE Novel human cDNA SEQ ID NO 332.  
 XX  
 KW ss; cancer; neurodegenerative disease; human.  
 XX  
 OS Homo sapiens.  
 XX  
 FN US2003104529-A1.  
 XX  
 PD 05-JUN-2003.  
 XX  
 PF 04-JAN-2002; 2002US-00037270.  
 XX  
 PR 21-JAN-2000; 2000US-00488725.  
 PR 25-APR-2000; 2000US-00552317.  
 PR 19-JUL-2000; 2000US-00620312.  
 XX

PA (ZHOU/) ZHOU P.  
 PA (TANG/) TANG Y T.  
 PA (LIUC/) LIU C.  
 PA (ASUN/) ASUNDI V.  
 PA (DRMA/) DRMANAC R T.  
 PI Zhou P, Tang YT, Liu C, Asundi V, Drmanac RT;  
 XX WPI; 2003-678194/64.  
 DR  
 XX New polynucleotide, useful for treating diseases e.g., cancer or  
 PT neurodegenerative diseases.  
 CC  
 PS Claim 1; SEQ ID NO 332; 99pp; English.  
 XX  
 CC The invention relates to a polynucleotide comprising a sequence given in  
 CC the specification, or its mature protein-coding portion, or its  
 CC complement. The polynucleotide is useful for treating diseases e.g.,  
 CC cancer or neurodegenerative diseases and many others listed in the  
 CC specification. The present sequence represents a novel human cDNA. Note:  
 CC The sequence data for this patent did not form part of the printed  
 CC specification but was obtained in electronic format directly from USPTO  
 CC at seqdata.uspto.gov/sequence.html?DocID=20030104529.  
 XX  
 SQ Sequence 6378 BP; 1823 A; 1420 C; 1619 G; 1515 T; 0 U; 1 Other;

Query Match 17.9%; Score 76; DB 8; Length 6378;  
 Best Local Similarity 59.8%; Pred. No. 1.7e-13;  
 Matches 146; Conservative 0; Mismatches 95; Indels 3; Gaps 1;  
 QY 115 GATAATCATGTTTCATCTGGAACACAGTCCGTGGGAGGTGTACAGGA---GACTGTGGT 171  
 Db 229 GGTAAATGGGAAAGCTTCTGAGAAAGTCCATGGGCCGATGTATGGGAGATGAATGTGT 288  
 QY 172 CCCGAGGAGTCCAGAGTCGGGCAGTGTGTTTTCATGTTGACGGGTGGACAAGTCA 231  
 Db 289 CCCGAGGATCCAAACGAGGGCTGTGTGTGTCTCATGTGGAGGATGACTACACTG 348  
 QY 232 CTGCTTAAGTCTGGTGAGCAACAGGCCTCCAAAGGAAGATTGTTTCCGAGTTTGT 291  
 Db 349 CATACTAAGTCTAAGCAGCCGCGAGAGACCCCAATAACCAAGCAGAGATTGTTTCAAAAGTTGC 408  
 QY 292 GACTGGCAGTGTGAGTCTTTTCAGTGGGAGGTTTCTGACTGGCAGTGTGTTGTT 351  
 Db 409 GATTGGCAAAAGAGTGTACACTGTGAGACTGGGAGCTTGGAAATCAGTGTACGCCGTG 468  
 QY 352 CCTT 355  
 Db 469 ATTT 472

RESULT 10  
 AAI34372  
 ID AAI34372 standard; DNA; 1956 BP.

AC AAI34372;  
 XX  
 DT 17-OCT-2001 (first entry)  
 XX  
 DE Probe #3058 used to measure gene expression in human placenta sample.  
 XX  
 KW Probe; microarray; human; placenta; antenatal diagnosis;  
 KW Genetic disorder; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157272-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US000663.  
 XX  
 PR 04-FEB-2000; 2000US-0180312P.  
 XX

PR 26-MAY-2000; 2000US-0207456P.  
 PR 30-JUN-2000; 2000US-00609408.  
 PR 03-AUG-2000; 2000US-00632366.  
 PR 21-SEP-2000; 2000US-0234687P.  
 PR 27-SEP-2000; 2000US-0236359P.  
 PR 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.  
 PA  
 XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2001-488897/53.

XX Human genome-derived single exon nucleic acid probes useful for analyzing  
 XX gene expression in human placenta.  
 XX Claim 25; SEQ ID NO 3058; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SENP).  
 CC The present sequence is one such probe. The probes are useful for  
 CC producing a microarray for predicting, measuring and displaying gene  
 CC expression in samples derived from human placenta. The probes are useful  
 CC for antenatal diagnosis of human genetic disorders

XX Sequence 1956 BP; 595 A; 370 C; 451 G; 540 T; 0 U; 0 Other;  
 Query Match 17.8%; Score 75.4; DB 4; Length 1956;  
 Best Local Similarity 62.0%; Pred. No. 1.6e-13;  
 Matches 13; Conservative 0; Mismatches 81; Indels 3; Gaps 1;  
 QY 138 AGTCCCTGGGGAAGTGTACAGGA---GACTGTGGTCCCGAGGAGTCCAGAGTCGGGC 194  
 Db 189 AGTCCATGGGCGGATGTATGGAGATGAATGTGGTCCCGAGGATCCAAACGAGGC 248  
 2Y 195 ACGTGTGGTGTTCATGTGACGGGTGGACAGTCACTGTCTTAATGTGGTGGAGAGCAA 254  
 Db 249 TGTGTGGTGTCTCATGTGGAGGATGGACTACACTGCATCTAACTGTAAGCAGGCCGA 308  
 2Y 255 CAGGCTCCAAAGAAAGTGTTCCTCGAGTTTGTGACTGGCAGAGTCACTCTTTCA 314  
 Db 309 GAGACCCAAATACCAGCAGATTTGTTCAAGTTTGGATTGGCAAAAGAGTTGTACGA 368  
 2Y 315 GTGGAGGTTTCTGACTGGCACCAGTGTGGTGTTCCTT 355  
 Db 369 CTGGAGACTGGGACCTTGAATCAGTGTGAGCCGTGATTT 409

## RESULT 11

AB28046  
 ID ABS28046 standard; DNA; 1956 BP.

XX ABS28046;  
 XX

XX 25-FEB-2003 (first entry)  
 XX

XX Human liver single exon probe, SEQ ID No 3036.  
 DE

XX Human; single exon nucleic acid probe; liver; cirrhosis;  
 KW hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;  
 KW coronary heart disease; ss.  
 XX

XX Homo sapiens.  
 XX

XX W0200157273-A2.  
 XX

XX 09-AUG-2001.  
 XX

XX 30-JAN-2001; 2001WO-US000664.  
 PF

XX 04-FEB-2000; 2000US-0180312P.  
 PR

XX 26-MAY-2000; 2000US-0207456P.  
 PR

XX 30-JUN-2000; 2000US-00609408.  
 PR

XX 03-AUG-2000; 2000US-00632366.  
 PR

PR 21-SEP-2000; 2000US-0234687P.  
 PR 27-SEP-2000; 2000US-0236359P.  
 PR 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.  
 PA  
 XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2001-488898/53.

XX Human genome-derived single exon nucleic acid probes useful for analyzing  
 XX gene expression in human adult liver.  
 XX Claim 1; SEQ ID NO 3036; 658pp; English.

XX The invention relates to a single exon nucleic acid probe (SENP) (I) for  
 CC measuring human gene expression in a sample derived from human adult  
 CC liver, comprising one of 13109 defined nucleotide sequences given in the  
 CC specification (or complements/fragments). The probe hybridizes at high  
 CC stringency to a nucleic acid molecule expressed in the human adult liver.  
 CC (I) may be used for predicting, measuring and displaying gene expression  
 CC in samples derived from human adult liver. The genes identified may be  
 CC involved in genetic liver diseases such as cirrhosis,  
 CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is  
 CC associated with coronary heart disease. ABS25011-ABS51005 represent human  
 CC liver single exon nucleic acid probes of the invention. Note: The  
 CC sequence information for this patent does not appear in the printed  
 CC specification but was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 1956 BP; 595 A; 370 C; 451 G; 540 T; 0 U; 0 Other;  
 Query Match 17.8%; Score 75.4; DB 4; Length 1956;  
 Best Local Similarity 62.0%; Pred. No. 1.6e-13;  
 Matches 137; Conservative 0; Mismatches 81; Indels 3; Gaps 1;

QY 138 AGTCCCTGGGGAAGTGTACAGGA---GACTGTGGTCCCGAGGAGTCCAGAGTCGGGC 194  
 Db 189 AGTCCATGGGCGGATGTATGGAGATGAATGTGGTCCCGAGGATCCAAACGAGGC 248  
 QY 195 AGTGTGGTGTTCATGTGACGGGTGGACAGTCACTGTCTTAATGTGGTGGAGAGCAA 254  
 Db 249 TGTGTGGTGTCTCATGTGGAGGATGGACTACACTGCATCTAACTGTAAGCAGGCCGA 308  
 QY 255 CAGGCTCCAAAGAAAGTGTTCCTCGAGTTTGTGACTGGCAGAGTCACTCTTTCA 314  
 Db 309 GAGACCCAAATACCAGCAGATTTGTTCAAGTTTGGATTGGCAAAAGAGTTGTACGA 368  
 QY 315 GTGGAGGTTTCTGACTGGCACCAGTGTGGTGTTCCTT 355  
 Db 369 CTGGAGACTGGGACCTTGAATCAGTGTGAGCCGTGATTT 409

## RESULT 12

AAF30188  
 ID AAF30188 standard; cDNA; 6373 BP.

XX AAF30188;  
 XX

XX 30-APR-2001 (first entry)  
 XX

XX Clone 21433858 encoding secreted protein SECP1.  
 DE

XX SECP1; secreted protein; human; diagnosis; gene therapy; lung cancer; ss.  
 XX

XX Homo sapiens.  
 XX

XX Key Location/Qualifiers  
 FT CDS 235..5001  
 FT sig\_peptide /\*tag= a  
 FT mat\_peptide /\*tag= b

FT 304..4998





Kekuda R, Leach MD, Li L, Miller CB, Patturajan M, Rieger DK;  
 Rothenberg ME, Shimkets RA, Smithson G, Spytek KA, Taupier RJ;  
 Vernet CM, Voss EZr, Zerhusen BD, Zhong M;  
 WPI; 2003-381625/36.  
 P-PSDB; ABR58459.  
 NOVX polypeptides and nucleic acids useful for diagnosing, preventing or  
 treating NOVX-associated disorders, e.g. diabetes, obesity, cancer or  
 dyslipidemia, and in chromosome mapping, tissue typing or  
 pharmacogenomics.  
 Claim 20; Page 263-264; 487pp; English.  
 The present invention relates to novel human NOV proteins and their  
 coding sequences (ACC72075-ACC72181 and ABR58363-ABR58469). The NOV  
 proteins are useful in manufacturing a medicament for treating a syndrome  
 associated with a human disease. The NOV proteins and coding sequences  
 may be used to diagnose, treat or prevent metabolic disorders such as  
 diabetes or obesity, infections, cachexia, cancer, neurodegenerative  
 disorders such as Alzheimer's disease or Parkinson's disease, immune  
 disorders, haematopoietic disorders and various dyslipidaemias  
 Sequence 6373 BP; 1836 A; 1385 C; 1603 G; 1548 T; 0 U; 1 Other;  
 Query Match 17.8%; Score 75.4; DB 7; Length 6373;  
 Best Local Similarity 62.0%; Pred. No. 2.7e-13;  
 Matches 137; Conservative 0; Mismatches 81; Indels 3; Gaps 1;  
 QY 138 AGGTCCGTGGGGAGGTGTACAGGA---GACTGTGTCCTCCGGAGGAGTCCAGAGTCGGGC 194  
 |||||  
 Db 216 AGGTCCATGGGGCCGCGATGTGGAGATGAATGTGTCCTCCGGAGCATCCAAACGAGGCG 275  
 QY 195 AGTGTGTGTTTCATGTTTCAGCGGTGGACAGTCACTCTCTAACTGTGTGTGAGACAA 254  
 |||||  
 Db 276 TGTGTGTGTGCTCATGTGGAGGATGGACTACATGTCATACTAACTGTAAAGAGGCCGA 335  
 QY 255 CAGCGCTCCAAAGGAAAGAGTTGTTTCGAGTTCGACTGGCACAGTGCACCTCTTTCA 314  
 |||||  
 Db 336 GAGACCAATAACACGAGAA TTGTTTCAAGTTTGGCATTGGCACAAAGATTGTACGA 395  
 QY 315 CTGGGAGTTTCTGACTGGCACCACCTGTCGTTGTTTCCTT 355  
 |||||  
 Db 396 CTGAGACTGGGACCTTGGAAATCAGTGTGAGCCCGTGATT 436  
 RESULT 15  
 ACC72168  
 ID ACC72168 standard; DNR; 6373 BP.  
 XX  
 AC ACC72168;  
 XX  
 DT 07-JUL-2003 (first entry)  
 XX  
 DE Human NOV47g coding sequence.  
 XX  
 KW Human; NOV; antidiabetic; anorectic; antibacterial; virucide;  
 XX immunomodulator; cytostatic; nootropic; neuroprotective;  
 KW antiparkinsonian; antilipaeamic; gene therapy; metabolic disorder;  
 KW diabetes; obesity; infection; cachexia; cancer; dyslipidaemia;  
 KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
 KW immune disorder; haematopoietic disorder; gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003029423-A2.  
 PD 10-APR-2003.  
 XX  
 PF 02-OCT-2002; 2002WO-US031358.  
 XX  
 PR 05-OCT-2001; 2001US-0326483P.  
 XX 02-OCT-2001; 2001US-0327342P.  
 PR

Job time : 150.635 secs

```
PR 09-OCT-2001; 2001US-0327917P.
PR 09-OCT-2001; 2001US-0328029P.
PR 09-OCT-2001; 2001US-0328044P.
PR 09-OCT-2001; 2001US-0328056P.
PR 12-OCT-2001; 2001US-0328849P.
PR 15-OCT-2001; 2001US-0329414P.
PR 17-OCT-2001; 2001US-0330142P.
PR 22-OCT-2001; 2001US-0341058P.
PR 24-OCT-2001; 2001US-0339266P.
PR 24-OCT-2001; 2001US-0343629P.
PR 29-OCT-2001; 2001US-0349575P.
PR 01-NOV-2001; 2001US-0346357P.
PR 12-APR-2002; 2002US-0371972P.
PR 12-APR-2002; 2002US-0371980P.
PR 17-APR-2002; 2002US-0373261P.
PR 19-APR-2002; 2002US-0373805P.
PR 23-APR-2002; 2002US-0374738P.
PR 16-MAY-2002; 2002US-0381101P.
PR 17-MAY-2002; 2002US-0381635P.
PR 29-MAY-2002; 2002US-0383830P.
PR 01-OCT-2002; 2002US-00262839.
XX
FA (CURA-) CURAGEN CORP.
XX
XX Alsobrook JP, Anderson DW, Boldog FL, Burgess CE, Catterton E;
PI Edinger SR, Ellerman K, Gerlach VL, Gorman L, Guo X, Ji W;
PI Kekuda R, Leach MD, Li L, Miller CE, Patturajan M, Rieger DK;
PI Rothenberg ME, Shinkens RA, Smithson G, Spytek KA, Taupier RJ;
PI Vernet CAM, Voss EZ, Zerhusen BD, Zhong M;
XX
DR WPI; 2003-381625/36.
DR P-PSDB; ABR58456.
XX
XX NOVX polypeptides and nucleic acids useful for diagnosing, preventing or
PT treating NOVX-associated disorders, e.g. diabetes, obesity, cancer or
PT dyslipidemia, and in chromosome mapping, tissue typing or
XX pharmacogenomics.
XX
PS Claim 20; Page 260-261; 487pp; English.
XX
XX The present invention relates to novel human NOV proteins and their
CC coding sequences (ACC72075-ACC72181 and ABR58363-ABR58469). The NOV
CC proteins are useful in manufacturing a medicament for treating a syndrome
CC associated with a human disease. The NOV proteins and coding sequences
CC may be used to diagnose, treat or prevent metabolic disorders such as
CC diabetes or obesity, infections, cachexia, cancer, neurodegenerative
CC disorders such as Alzheimer's disease or Parkinson's disease, immune
CC disorders, haematopoietic disorders and various dyslipidaemias
XX
XX Sequence 6373 BP; 1836 A; 1385 C; 1603 G; 1548 T; 0 U; 1 Other;
SQ
Query Match 17.8%; Score 75.4; DB 7; Length 6373;
Beat Local Similarity 62.0%; Pred. No. 2.7e-13;
Matches 137; Conservative 0; Mismatches 81; Indels 3; Gaps 1;
OY 138 AGGTCCGTGGGGAAGTCTACAGG--GACTGTGGTCCCGAGGAGTCCAGAGTCGGGC 194
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
216 AGGTCCATGGGCGCATGTATGGGAGATGAATGTGGTCCCGAGGCGATCCAAACGAGGGC 275
OY 195 AGTGTGGTGTCTTTCATGTGTGACGGGTGACCAAGTCACCTGTCTAACTGTGTGAGAGCAA 254
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
276 TGTGTGGTGTGTCTCATGTGGAGGATGACTACACTGCATCTAACTGAAGCAGGCCGA 335
OY 255 CAGGCCTCAAGAGGAAGAAGTGTTCGGAGTTTGTGACTGGCAGCAGTACCTCTTTCA 314
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
336 GAGACCCCAATACCAAGCAGAAATGTTTCAAAGTTTGGGATTGGCACAAGAGTTGTACGA 395
OY 315 GTGGGAGGTTTCTGACTGCACACTGTGTGCTTTCCTT 355
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
396 CTGGAGACTGGGACCTTGGATCACTGTACGCCCGTGATTT 436
```

Search completed: February 23, 2004, 12:24:19

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

M nucleic - nucleic search, using sw model

un on: February 23, 2004, 11:46:56 / Search time 968.64 Seconds  
(without alignments)  
13071.481 Million cell updates/sec

title: US-10-022-710-5\_COPY\_1\_424  
effect score: 424  
equence: 1 agttttcaagagcaact.....gcattggactgcagaccgga 424

coring table: IDENTITY\_NUC  
Gapop 10.0 / Gapext 1.0

searched: 27513289 seqs, 14931090276 residues

total number of hits satisfying chosen parameters: 55026578

inimum DB seq length: 0  
aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

atabase :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hrc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pin:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vri:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

result	No.	Score	Match	Length	ID	Description
1	422.4	99.6	675	14	CF905312	CF905312 A0433A03-
2	294.4	69.4	626	10	BB655661	BB655661 BB655661
3	294.4	69.4	685	13	BU058391	BU058391 UI-M-FOO-
4	294.4	69.4	3242	11	AK083798	AK083798 Mus muscu

5	294.4	69.4	3897	11	AK051714	Mus muscu
6	272	64.2	686	10	BB656175	BB656175
7	213.6	50.4	474	10	BB656509	BB656509
8	197.2	46.5	632	10	BB647268	BB647268
9	196.6	46.4	2266	11	AK047508	Mus muscu
10	171.6	40.5	411	13	BY247109	BY247109
11	109.8	28.3	456	10	BF706397	BF706397
12	104.2	24.6	557	11	AK052119	Mus muscu
13	104.2	24.6	565	10	BB472930	BB472930
14	104.2	24.6	567	10	BB658510	BB658510
15	102.2	24.1	676	14	CF531936	CF531936
16	94.4	22.3	704	14	CA361329	UI-M-GHO-
17	79	18.6	983	29	CNS0368U	AL229927 Tetraodon
18	77.8	18.3	1201	9	AL531002	AL531002 AL531002
19	75.4	17.8	1148	11	AK080721	Mus muscu
20	75	17.7	659	10	BB657304	BB657304
21	67.6	15.9	1201	9	AL531003	AL531003 AL531003
22	65.2	15.4	620	10	BB656281	BB656281
23	64.8	15.3	888	14	CB192844	AGENCOURT
24	62	14.6	410	13	BY004406	BY004406
25	48.6	11.5	392	13	BY004152	BY004152
26	41	9.7	821	14	CK018416	CK018416
27	39.4	9.3	880	14	CK016846	AGENCOURT
28	39.2	9.2	433	14	R83422	YB82f01.r1
29	38.8	9.2	458	10	BB728992	BB728992
30	38	9.0	575	12	BG556716	EM_40.F1
31	37.4	8.8	463	10	AW258602	uq37c10.Y
32	36.6	8.6	869	14	CK015752	AGENCOURT
33	36.4	8.6	478	10	BB284592	BB284592
34	36.4	8.6	1101	29	CNS00KK2	AL077673 Drosophil
35	36.2	8.5	736	14	CB723795	UI-M-FY0-
36	36.2	8.5	755	14	CB246260	UI-M-FOO-
37	36.2	8.5	765	14	CA327230	UI-M-FY0-
38	36	8.5	568	10	AW422427	AW422427 f162e12.Y
39	36	8.5	1201	13	EX446359	EX446359
40	35.8	8.4	524	12	BG093822	mab44f11.
41	35.8	8.4	532	14	CA890203	B0157G12-
42	35.8	8.4	581	12	BM021611	BM021611 id42h07.X
43	35.8	8.4	584	12	BM021692	id44b01.X
44	35.8	8.4	632	12	BI964985	id34f08.X
45	35.8	8.4	650	14	CF539400	UI-M-GHO-

ALIGNMENTS

RESULT 1	CF905312	675 bp	mRNA	linear	EST 04-NOV-2003
LOCUS	A0433A03-5	NIA Mouse Osteoblast	cDNA Library	(long 1)	Mus musculus
DEFINITION	CDNA clone NIA:A0433A03	IMAGE:30741122	5'	mRNA sequence.	
ACCESSION	CF905312				
VERSION	CF905312.1	GI:38172361			
KEYWORDS	EST.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	1 (bases 1 to 675)				
AUTHORS	Piao, Y., KO, N.T., Lim, M.K. and KO, M.S.H.				
TITLE	Construction of long-transcript enriched cDNA libraries from				
	submicrogram amounts of total RNAs by a universal PCR amplification				
	method				
JOURNAL	Genome Res. 11 (9), 1553-1558 (2001)				
MEDLINE	21429098				
PUBMED	11544199				
COMMENT	Contact: Dawood B. Dudekula				
	Laboratory of Genetics				
	National Institute on Aging/National Institutes of Health				
	333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA				
	Email: cdna@slsun.grc.nia.nih.gov				
	Plate: AC433 row: A column: 03				
	Seq primer: MJ3 Reverse				

High quality sequence stop: 675  
POLYA=No.

# FEATURES

source  
1. . 675  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C3H/He mice"  
/db\_xref="nigEST:A0433A03-5"  
/db\_xref="taxon:10090"  
/clone="NIA:A0433A03 IMAGE:30741122"  
/dev\_stage="KUSA/A1 cells"  
/lab\_host="DH10B"  
/clone\_lib="NIA Mouse Osteoblast cDNA Library (Long 1)"  
/note="vector: pCMV-Sport6 (Invitrogen); Site 1: SalI;  
Site 2: NotI; Mouse cDNA project by the Laboratory of  
Genetics, National Institute on Aging (NIA), Intramural  
Research Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA).  
This is a long-transcript enriched cDNA library [Ref.  
Genome Res. 11: 1553-1558 (2001) [PMID: 11544199]]. Total  
RNAs were obtained from Dr. Akihiro Umezawa (Keio  
University School of Medicine, Japan). Double-stranded  
cDNAs were synthesized with an oligo(dT) primer  
[Invitrogen]:  
5'-pGATGCTCTAGATCGAGCGCGCCCTTTT-3' from  
2.1 ug of total RNA, treated with T4 DNA polymerase, and  
purified by ethanol-precipitation. The cDNAs were ligated  
to Lona-linker LI-Sal4, purified by phenol/chloroform, and  
separated from free linkers by Centricon 100. Then, the  
cDNAs were amplified by long-range high fidelity PCR using  
Ex Taq polymerase (Takara) with a primer Sal4-S. The  
products were purified by phenol/chloroform and Centricon  
100. The cDNAs were digested with SalI and NotI enzymes  
and cloned into SalI/NotI site of pCMV-SPORT6 plasmid  
vector. The DH10B E. coli host was transformed with the  
ligation mixture by the standard chemical method. The  
average insert size is about 3.0 kb. The library was  
constructed by Yulan Piao."

## ORIGIN

Query Match 99.6%; Score 422.4; DB 14; Length 675;  
Best Local Similarity 99.8%; Pred. No. 4.6e-121;  
Matches 423; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 ATGTTTCAAAGAGCAACCTTAACAGTCACCTTGGTATGGAGGAGCATGAGGAAGCTC 60  
DB 138 ATGTTTCAAAGAGCAACCTTAACAGTCACCTTGGTATGGAGGAGCATGAGGAAGCTC 197  
QY 61 TTTCTATTGCTTCTCTCTGCTCCATGCGAGCTCATTTGGAGGCAAAAGGATAAT 120  
DB 198 TTTCTATTGCTTCTCTCTGCTCCATGCGAGCTCATTTGGAGGCAAAAGGATAAT 257  
QY 121 CAGTTTCATCTGAAACACGAGTCCGTGGGGAGGTTGTACAGGAGACTGTGTCCTCCGGAGGA 180  
DB 258 CAGTTTCATCTGAAACACGAGTCCGTGGGGAGGTTGTACAGGAGACTGTGTCCTCCGGAGGA 317  
QY 181 GTCCAGAGTCGGGACAGTGTGGTGTTCATGTTGACGGGTGGACAGTCACTGTCTTAAC 240  
DB 318 GTCCAGAGTCGGGACAGTGTGGTGTTCATGTTGACGGGTGGACAGTCACTGTCTTAAC 377  
QY 241 TGTGTCGAGAGCAACAGGCTCCAAAGGAAGAGTGTTCCTGAGTTTGTGACTGGCAC 300  
DB 378 TGTGTCGAGAGCAACAGGCTCCAAAGGAAGAGTGTTCCTGAGTTTGTGACTGGCAC 437  
QY 301 AGTGACCTTTTCAGTGGAGGTTTTCAGTCGGACCACTGTGCTGTTCCTTACGCT 360  
DB 438 AGTGACCTTTTCAGTGGAGGTTTTCAGTCGGACCACTGTGCTGTTCCTTACGCT 497  
QY 361 CGCGTGAAGTCAAGCTTCGAGTGTGACGGCTGACGGCTGACGATGACATGACGAC 420  
DB 498 CGCGTGAAGTCAAGCTTCGAGTGTGACGGCTGACGGCTGACGATGACATGACGAC 557  
QY 421 CGGA 424  
|||||

Db 558 CGGA 561

RESULT 2  
LOCUS BB655661  
DEFINITION BB655661 RIKEN full-length enriched, 12 days embryo spinal ganglion  
Mus musculus cDNA clone D13001122 5', mRNA sequence.  
BB655661  
ACCESSION BB655661  
VERSION BB655661.1 GI:16489489  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (Bases 1 to 626)  
Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T.,  
Hara, A., Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J.,  
Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K.,  
Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K.,  
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,  
Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M. and Hayashizaki, Y.  
RIKEN Mouse ESTs (Arakawa, T., et al. 2001)  
Unpublished (2001)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,  
Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T.,  
Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A.  
and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P.,  
Sugahara, Y. and Hayashizaki, Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I.,  
Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and  
Hayashizaki, Y.  
Computational Analysis of Full-length Mouse cDNAs Compared with  
Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)  
Please visit our web site (http://genome.gsc.riken.go.jp) for  
further details.  
e mouse tissues.  
FEATURES  
source  
1. . 626  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="D13001122"  
/tissue\_type="spinal ganglion"  
/dev\_stage="12 days embryo"  
/lab\_host="DH10B"  
/clone\_lib="RIKEN full-length enriched, 12 days embryo  
spinal ganglion"  
/note="Site 1: SalI; Site 2: BamHI; cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5',  
GAGGAGAGAGCGCGCGCACTCGAGTTTCTTTTCTTTTNN 3'], cDNA was prepared by using trihalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5',  
GAGAGAGATCTCGAGTTTAAATTAATCCGCCGCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

## RIGIN

Query Match 69.4%; Score 294.4; DB 10; Length 626;  
Best Local Similarity 80.9%; Pred. No. 6.6e-81;  
Matches 343; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

Y 1 ATGTTTCCAAAGAGCAACTAACAGTCACTTCTGCTGGTATGAGGAGCATGAGAGCTC 60  
b 160 ATGTTTCTGAGAGTGATCTGGCTGTCTACTCACTGGGTTTCCAGGAGCATGAGAGCTC 219  
Y 61 TTTCATATGCTTTCTCTCTTCTGCTGCCATGCACTCAATTTGGAAGGCAAAAGGATAAT 120  
b 220 TTCTAGTGTCTTCTCTCTCTGCTGCCAGCACTCAATTTGGAAGGCAAGGACAAAT 279  
Y 121 CAGTTTCATCTGGAACACAGTCCGCTGGGGAAGGTGTACAGAGACTGTGCTCCCGAGGA 180  
b 280 CAGTTTCCTTGGAACACCGCTCCCTGGGGAAGGTGTGAGGTGTGCTGGGACCGAGCGGA 339  
Y 181 GTCCAGAGTCGGGAGTGTGGTGTCTTTCATGTGACGGGTGGACAAAGTCACTGTCTAAC 240  
b 340 GCCCAAGTGTGTGTATGTGTCTTTCATAGAGAGGTGGACAGTCCCATGTCCGAC 399  
Y 241 TGTGTGTAGAGCAACAGCGCTCCAAAGAAAGAGTGTCTTCCAGGTTTGTGACTGGCAC 300  
b 400 TGTGTGTAGAGCAACAGCGCTCCAAAGAAAGAGTGTCTTCCGCGTGTGTGACTGGCAC 459  
Y 301 AGTGACCTTTCAGTGGAGGTCTTCTGCTGGGACCACTGTGCTGTGTTTCTTACGCT 360  
b 460 AGTGACCTTTCAGTGGAGGTCTTCCAGTGGCATGCTGCTGTGTTTCTTGGTCTTGGGCCC 519  
Y 361 CGCGTGAAGTCAAGCCTCGGACTGCAGAGTGTGTGAGCGGTCAAGTGTGACTGCAGCAC 420  
b 520 CAAGGTGAGCCCGAGCGCTCGGCTGTGGAATGTGTGACTGCACAGCAGGCTTACAGCAC 579  
Y 421 CGGA 424  
b 580 CGGA 583

RESULT 3  
U058391  
OCUS  
EPIFINITION  
UI-M-F00-cae-h-19-0-UI-r1 NIH\_BMAP\_F00 Mus musculus cDNA clone  
IMAGE:6410178 5', mRNA sequence.  
CCESION  
U058391  
ERSON  
U058391.1 GI:22498667  
EYWORDS  
EST.  
Mus musculus (house mouse)  
ORIGIN  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 685)  
NIH-MGC <http://mgi.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Dr. Jim Lin, University of Iowa  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

FEATURES  
source

Seq primer: pYX-5.  
Location/Qualifiers  
1..685  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="CS7BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6410178"  
/tissue\_type="whole brain"  
/dev\_stage="embryo 12.5dpc"  
/lab\_host="DH10B (T1 phage resistant)"  
/clone\_lib="NIH BMAP\_F00"  
/notes="Organ: Brain; Vector: pYX-Asc; Site\_1: EcoR I; Site\_2: Not I; The library was constructed according Bonaudo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction. ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is TGAAGAGCC. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

## ORIGIN

Query Match 69.4%; Score 294.4; DB 13; Length 685;  
Best Local Similarity 80.9%; Pred. No. 6.9e-81;  
Matches 343; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 1 ATGTTTCCAAAGAGCAACCTAACAGTCACTTCTGCTGGTATGAGGAGCATGAGAGCTC 60  
Db 162 ATGTTTCTGAGGAGTGATCTGGCTGTCACTCACTGGGTTTCCAGGAGCATGAGAGCTC 221  
QY 61 TTTCATATGCTTTCTCTCTTGTGTCTCCCATGCAAGTCAATTTGGAAGGCAAAAGGATAAT 120  
Db 222 TTTCATATGCTTTCTCTCTCTGCTGTCTCCCAAGCAGCTCAATTTGGAAGGCAAGGACAAAT 281  
QY 121 CAGTTTCATCTGGAACACAGGTCCGTGGGGAAGGTGTACAGGAGACTGTGCTCCCGAGGA 180  
Db 282 CAGTTTCCTCTGGAACACCGGTCCCTGGGGAAGGTGTGACGAGTGTGCTGGGACCGGA 341  
QY 181 GTCAGAGTCGGGCACTGTGTGTTTTCATGTGACGGGTGGACAAAGTCACTGTCTTAAC 240  
Db 342 GCCCAAGT 401  
QY 241 TGTGTTGAGAGCAACAGGCTCCAAAGGAAAGAGTGTGTTTCCGAGTTTGTGACTGGCAC 300  
Db 402 TGTGATGAGAGCAGCCAACTCCAAAGGAAAGAGTGTGTTTCCGCGTGTGTGACTGGCAC 461  
QY 301 AGTGACCTCTTTCAGTGGGAGGTCTGACTGGCACCACCTGTGCTGTGCTTCTTACGCT 360  
Db 462 AGTGACCTGTTCAGTGGGAGGTTCCTGACTGGCATCGCTGTGCTTCTTCTTCTTCTTCTTCT 521  
QY 361 CGCGTGAAGTCAAGCCTCGGAGCTGCGAGAGTGTGTGACCGGCTCAGCATGGAATGAGAGCAC 420  
Db 522 CAAGGTGAGCCCGAGCGCTCGGCTGTGGAATGTGTGACTGTGACAGCAGCAGCTCTACAGCAC 581  
QY 421 CGGA 424  
Db 582 CGGA 585

RESULT 4  
AK083798

LOCUS  
DEFINITION  
AK083798 3242 bp mRNA linear HTC 20-SEP-2003  
Mus musculus 12 days embryo spinal ganglion cDNA, RIKEN full-length  
enriched library, clone:DJ3001122 product:weakly similar to  
HYPOTHETICAL PROTEIN DJ1110N13.1 IN CHROMOSOME 7 (FRAGMENT) [Homo  
sapiens], full insert sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
AK083798.1 GI:26101501  
HTC; CAP trapper.  
Mus musculus  
Mus musculus (house mouse)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
99279253  
PUBMED  
10349636  
REFERENCE  
2  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
20499374  
PUBMED  
11042159  
REFERENCE  
3  
Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P.,  
Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujikawa, S., Inoue, K., Togawa, Y., Izawa, M., Ohata, E., Watahiki, M.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
20530913  
PUBMED  
11076861  
REFERENCE  
4  
The RIKEN Genome Exploration Research Group Phase II Team and the  
PANTOM Consortium.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)  
5  
The FANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
6 (bases 1 to 3242)  
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,  
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,  
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,  
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,  
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, N.,  
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,  
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,  
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,  
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
Muramatsu, M. and Hayashizaki, Y.  
Direct Submission  
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,  
Fax: 81-45-503-9216)  
COMMENT  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.go.jp/  
URL: http://fantom.gsc.riken.go.jp/.  
FEATURES  
Location/Qualifiers  
1..3242  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="PANTOM\_DB:DJ3001122"  
/db\_xref="MGI:2419369"  
/db\_xref="taxon:10090"  
/clone="DJ3001122"  
/tissue type="spinal ganglion"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev stage="12 days embryo"  
misc\_feature  
1..3242  
/note="weakly similar to HYPOTHETICAL PROTEIN DJ1110N13.1  
IN CHROMOSOME 7 (FRAGMENT) [Homo sapiens] (SPTR:O43376,  
evidence: PASTY, 53.6%ID, 99.4%length, match=540)"  
ORIGIN  
Query Match 69.4%; Score 294.4; DB 11; Length 3242;  
Best Local Similarity 80.9%; Pred. No. 1.4e-80;  
Matches 343; Conservative 0; Mismatches 81; Indels 0; Gaps 0;  
QY 1 ATGTTTCAAAGAGCAACCTAACAGTCACTTCTGGGTATGAGGAGCATGAGGAAGCTC 60  
Db 160 ATGTTTCTGAGGAGTGATCTGGCTCACTCACTGGGTTTCCAGGAGCATGAGGAAGCTC 219  
QY 61 TTTCATTGCTTTCTCTCTCTGCTGCCATCACTCACTTTTGAAGCCAAAGGATAAT 120  
Db 220 TTTCATTGCTTTCTCTCTCTGCTGCCATCACTCACTTTTGAAGCCAAAGGATAAT 279  
QY 121 CAGTTTCATCTGGAACACAGGTCCTGGGGAGGTTGTACAGGAGACTGTGGTCCCGAGGA 180  
Db 280 CAGTTTCCTCGAAGAACCGGTCCTGGGGAGGTTGTACAGGAGACTGTGGTCCCGAGGA 339  
QY 181 GTCCAGAGTCGGGAGTGTTGGTGTTCATGTTTACGGGTGGACAGTCACTCTTAAC 240  
Db 340 GCCCAAGTCGTGTGTATGTGTCTTTCACATGAAGGGTGGAGTCCCATGTCCGAA 399  
QY 241 TGTGTTGAGAGCAACAGGCTCCAAAGAAAGAGTTGTTTCCGAGTTTGTGACTGGCAC 300  
Db 400 TGTGATGAGAGCAGCCAACTCCAAAGAAAGAGTGTTCGCGGTGTGTGACTGGCAC 459  
QY 301 AGTGACCTCTTTCAGTGGAGGTTTCTGACTGGACCACTGTGTGCTTGTCTTACGCT 360  
Db 460 AGTGACCTGTTCAGTGGAGGTTTCCGACTGGCATCGCTTCTGTTCTCTGGGGCC 519  
QY 361 CGCGGTGAAGTCAAGCCTCGGACTCGAGGTGTGTGCGGTGAGCATGAGTGGACTCGAGCAC 420  
Db 520 CAGGTGAGCCAGGCTCGGCTGTGGAATGTGTGACTGACAGCAGCGGTCTACAGCAC 579  
QY 421 CGGA 424  
Db 580 CGGA 583  
RESULT 5  
AK051714 3897 bp mRNA linear HTC 20-SEP-2003  
LOCUS  
DEFINITION  
Mus musculus 12 days embryo spinal ganglion cDNA, RIKEN full-length  
enriched library, clone:DJ30067103 product:weakly similar to  
HYPOTHETICAL PROTEIN DJ1110N13.1 IN CHROMOSOME 7 (FRAGMENT) [Homo  
sapiens], full insert sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
AK051714  
AK051714.1 GI:26342139  
HTC; CAP trapper.  
Mus musculus (house mouse)  
Mus musculus





QY 421 CGGA 424  
 ||||  
 Db 594 CGGA 597

## RESULT 6

BB656175

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BB656175 686 bp mRNA linear EST 26-OCT-2001  
 BB656175 RIKEN full-length enriched, 12 days embryo spinal ganglion  
 Mus musculus cDNA clone D130067I03 5', mRNA sequence.

BB656175

BB656175.1 GI:16490003

EST.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 686)

Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T.,  
 Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K.,  
 Ohno, H., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K.,  
 Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
 Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,  
 Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M. and Hayashizaki, Y.  
 RIKEN Mouse ESTs (Arakawa, T., et al. 2001)  
 Unpublished (2001)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsr.riken.go.jp,

URL: http://genome.gsc.riken.go.jp/

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

Wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,

Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T.,

Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A.

and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res.

10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P.,

Sugahara, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I.,

Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and

Hayashizaki, Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with

Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for

further details.

e mouse tissues.

## FEATURES

Location/Qualifiers

1. 686

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"

/clone="D130067I03"

/cissue\_type="spinal ganglion"

/dev\_stage="12 days embryo"

/lab\_host="DH10B"

/clone\_lib="RIKEN full-length enriched, 12 days embryo

spinal ganglion"

/note="Site 1: Sall; Site 2: BamHI; cDNA library was  
 prepared and sequenced in Mouse Genome Encyclopedia  
 Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in  
 RIKEN, Division of Experimental Animal Research in Riken  
 contributed to prepare mouse tissues. 1st strand cDNA was  
 primed with a primer [5'  
 GAGAGAGAGCGCGCAACTCGAGTGTGTTTTTTTTTTN 3']. cDNA was  
 prepared by using trehalose thermo-activated reverse  
 transcriptase and subsequently enriched for full-length by  
 cap-trapper. Second strand cDNA was prepared with the  
 primer adapter of sequence [5'  
 GAGAGAGAGATCTCGAGTTAATAATAATCCCCCCCCCC 3']. cDNA  
 was cleaved with BamHI and XhoI. Vector: a modified  
 pBluescript KS(+) after bulk excision from Lambda FLC 1."

## ORIGIN

Query Match 64.2%; Score 272; DB 10; Length 686;  
 Best Local Similarity 78.4%; Pred. No. 7, 4e-74;  
 Matches 323; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 1 ATGTTTCAAAGAGCAACTACACTCTGCTGGGTATGGAGGAGCATGAGGAAGCTC 60

Db 174 ATGTTTCTGAGGAGTGATCTGCTCTCACTCTGGTTCACGAGCATGAGGAAGCTC 233

QY 61 TTTCTATTGCTTTCTCTCTGCTGCCATCGAGCTCAATTTGGAGGCAAAAGGATAAT 120

Db 234 TTTCTAGTCTTTCTCTCTGCTGCCATCGAGCTCAATTTGGAGGCAAAAGGATAAT 293

QY 121 CAGTTTCATCTGAAACCAAGGTCCTGGGGAGAGGTACAGGAGAGCTGTGTCCCGAGGA 180

Db 294 CAGTTTCCTCGAAAACCGTCCCTGGGGAGAGGTGTGCAGGTGACTCGGACCGAGCGGA 353

QY 181 GTCCAGAGTCGGGAGTGTGGTGTTCATGTTGACGGGTGGAGAGTCACTGTCTTAAC 240

Db 354 GCCCAAGTCGTGTGTATGGTGTCTTTCACATAGAAGGGTGGAGAGTCCCATGTCTGAAC 413

QY 241 TGTGTTGAGAGCAACAGGCTCCAAAGGAAAGAGTGTTCCTCGAGTGTGTGACTGGCAC 300

Db 414 TGTGATGAGAGCAGCAACCTCCAAAGGAAAGAGTGTTCCTCGGCTGTGTGACTGGCAC 473

QY 301 AGTGACCTTTTCAGTGGAGGTTTCTGACTGGGACCACTGTGTGCTTGTTCCTTACGCT 360

Db 474 AGTGACCTGTTTCCAGGGGAGGTTTCCGACTGGCATGCTGCTTTTGTGTTCTCTGGGNC 533

QY 361 CGCGGTGAGTCAAGCTCGGACTCGAGAGTGTGTGACGCTCAGCATGGAC 412

Db 534 CCAGGGAGCCCGAGGCTTGGGCTGGGAGATGGGNGACTGCACAGCCCGGCG 585

## RESULT 7

BB656509

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BB656509 474 bp mRNA linear EST 26-OCT-2001  
 BB656509 RIKEN full-length enriched, 12 days embryo spinal ganglion  
 Mus musculus cDNA clone D130033M03 5', mRNA sequence.

BB656509

BB656509.1 GI:16490337

EST.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 474)

Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T.,  
 Hara, A., Hiramoto, K., Hori, P., Ishii, Y., Ito, M., Kawai, J.,  
 Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K.,  
 Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K.,  
 Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
 Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,  
 Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M. and Hayashizaki, Y.  
 RIKEN Mouse ESTs (Arakawa, T., et al. 2001)  
 Unpublished (2001)  
 Contact: Yoshihide Hayashizaki



FEATURES  
source

## Location/Qualifiers

```

1. .632
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/sex="mixed"
/tissue_type="cerebellum"
/dev_stage="10 days neonate"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 10 days neonate cerebellum"
/notes="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGGAGAGAGGATCCAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was prepared by using trisphosphate thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGGAGAGATCTCGAGTTAATTAAATTAATCCCTCCCCC Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"

```

## ORIGIN

```

Query Match 45.5%; Score 197.2; DB 10; Length 632;
Best Local Similarity .80.0%; Pred.No.2.1e-50;
Matches 232; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

Qy 135 ACCAGGTCCTGGGGAAGGTACAGGAGCTGTGGTCCCGAGGAGTCAGAGTCGGGC 194
Db 247 AGCAGGTCCTGGGGAAGGTGTGAGGTGACTGCGGACGAGCGGAGCCAAAGTCGTGC 306
Qy 195 AGTGTGGTGTTCATGTGACGGGTGACAGTCACCTGCTCACTGTGTGTGAGACCA 254
Db 307 TGTATGGTGCTTTCACATAGAGGTTGACAGTCCCATGTCGAACTGTGTATGAGACG 366
Qy 255 CAGGCTCCAAAGGAAGAAGTGTGTCGAGTTGTGACTGGCACAGTGACCTCTTTCA 314
Db 367 CCAACTCCAAAGGAAGAAGCTGTTCGCGGTGTGTGACTGGCACAGTGACCTGTCCA 426
Qy 315 GTGGAGGTTTCTGACTGGCACCACTGTGTGCTGTTCCTTACGCTCGCGGTGAAGTCAA 374
Db 427 GTGGAGGTTTCCGACTGGCATCGTGTGCTTCTGGTTCCTGGGCCCCAAGTGAGCCAG 486
Qy 375 GCTCGGACTCGAGAGTGTGTGACGGCTCAGCATGGAGCTGCAGCACCGGA 424
Db 487 GCTCGGCTGGGGAATGTGTGACTGCACACAGCGGTCTACGACCCGGA 536

```

## RESULT 9

```

AK047508
LOCUS AK047508 2266 bp mRNA linear HTC 20-SEP-2003
DEFINITION Mus musculus 10 days neonate cerebellum cDNA, RIKEN full-length enriched library, clone:B930082A18 product:weakly similar to HYPOTHETICAL PROTEIN DUIL0N13.1 IN CHROMOSOME 7 (FRAGMENT) [Homo sapiens], full insert sequence.

```

## ACCESSION AK047508.1 GI:26092224

## VERSION AK047508

## KEYWORDS HTC; CAP trapper.

## SOURCE Mus musculus (house mouse)

## ORGANISM

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

```

## REFERENCE 1

## AUTHORS

```

TITLE Carninci,P. and Hayashizaki,Y.
JOURNAL High-efficiency full-length cDNA cloning
MEDLINE Meth. Enzymol. 303, 19-44 (1999)
PUBMED 99279253
PUBMED 10349636

```

## REFERENCE

## AUTHORS

```

TITLE Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
JOURNAL Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
MEDLINE Normalization and subtraction of cap-trapper-selected cDNAs to
PUBMED prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)

```

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

```

TITLE Shibata,K., Itoh,M., Aizawa,K., Nagaoaka,S., Sasaki,N., Carninci,P.,
JOURNAL Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,
MEDLINE Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
PUBMED Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Matakaki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (KISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

```

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

```

TITLE The RIKEN Genome Exploration Research Group Phase II Team and the
JOURNAL FANTOM Consortium.
MEDLINE Functional annotation of a full-length mouse cDNA collection
PUBMED Nature 409, 685-690 (2001)

```

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

```

TITLE The FANTOM Consortium and the RIKEN Genome Exploration Research
JOURNAL Group Phase I & II Team.
MEDLINE Analysis of the mouse transcriptome based on functional annotation
PUBMED of 60,770 full-length cDNAs
REFERENCE Nature 420, 563-573 (2002)

```

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

```

TITLE Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
JOURNAL Fukuda,S., Furuno,M., Kanagaki,T., Hara,A., Hashizume,W.,
MEDLINE Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
PUBMED Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
Kato,H., Kawai,O., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)

```

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## COMMENT

```

cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL:http://genome.gsc.riken.go.jp/
URL:http://fantom.gsc.riken.go.jp/
Location/Qualifiers
1. .2266
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:B930082A18"
/db_xref="MGI:2413115"

```

## FEATURES

## source

```

/db_xref="taxon:10090"
/clone="B930082A18"
/issue_type="cerebellum"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="10 days neonate"
misc_feature
1..2266
/notes="weakly similar to HYPOPHYSICAL PROTEIN D1110N13.1
IN CHROMOSOME 7 (FRAGMENT) [Homo sapiens] (SPTRI043376,
evidence: FASTY, 53.6%id, 99.4%length, match=540)"

>RGIN
Query Match 46.4%; Score 196.6; DB 11; Length 2266;
Best Local Similarity 79.7%; Pred. No. 6e-50;
Matches 232; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

>Y
134 ACCAGGTCGGTGGGAAGGTGTACAGGAGATGTGGTCCCGAGGAGTCCAGATCGGG 193
251 AGCCAGGTCCTCGGGAAGGTGTGCGAGGTGACTGCGGACGAGCGGAGCCCAAGTCGTG 310

>Y
194 CAGTGTGGTGTTCATGTGTGCGGTGACAGAGTCACTCTTAACGTGTGTGTGAGACA 253
311 CTGTATGGTGTTCACATAGAGGAGGTGACGAGTCCCATGTGCAACTGTGTGAGAGACA 370

>Y
254 ACAGGCTCCAAAGGAAAGTGTGTTCCGAGTGTGACTGCGACAGTGAACCTTTTC 313
371 GCCAACCCTCAAAGGAAGAGCTGTTTCGGGTGTGTGACTGGCACAGTGAACCTTTCC 430

>Y
314 AGTGGAGGTTTCTGACTGGCACCACTGTGTGTTCTTACCGTTCGGGTGGAAGTCA 373
431 AGTGGAGGTTTCCGACTGGCATCGTGTCTTCTGCTTCTCGGCGCCCAAGGTGAGCCCA 490

>Y
374 AGCCTCGGACTCGAGAGTGTGTGACGGCTCAGCATGGACTCGAGCACCGGA 424
491 GGCTCGGCTGTGGATGTGTGACTGACAGCACAGCGTCTACAGACCGGA 541

```

```

RESULT 10
>Y247109
>LOCUS
>DEFINITION
BY247109 RIKEN full-length enriched, visual cortex Mus musculus
CDNA clone K230308M07 5', mRNA sequence.
BY247109
>CESSION
BY247109.1 GI:26428621
>KEYWORDS
EST.
>SOURCE
Mus musculus (house mouse)
>ORGANISM
Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 411)
Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
Nikaido,I., Oatso,N., Saito,R., Suzuki,H., Yamanaka,I.,
Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Bruscia,V.,
Chochia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
Kurochkin,I.V., Lee,Y., Lehar,B., Lyons,P.A., Maglott,D.R.,
Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
Verardo,R., Wagner,L., Wahlstedt,C., Wang,Y., Waranabe,Y.,
Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,

```

```

Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
Shingawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
Rogers,J., Birney,E. and Hayashizaki,Y.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
JOURNAL
2354683
MEDLINE
12466851
PUBMED
COMMENT
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gs.riken.go.jp/
URL:http://genome.gsc.riken.go.jp/
Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,
Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,
Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,
Onno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,
Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and
Hayashizaki,Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Tissues were provided by Michela Pagiolini and Takao K. Hensch (
Laboratory for Neuronal Circuit Development Brain Science Institute
RIKEN 2-1 Hirosewa,Wako-shi,Saitama 351-0198 Japan ) whose
assistance we gratefully acknowledge. Please visit our web site
(http://genome.gsc.riken.go.jp) for further details.
FEATURES
Location/Qualifiers
source
1..411
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="CS7BL/6J"
/clone="K230308M07"
/tissue_type="visual cortex"
/clone_lib="RIKEN full-length enriched, visual cortex"
ORIGIN
Query Match 40.5%; Score 171.6; DB 13; Length 411;
Best Local Similarity 81.8%; Pred. No. 1.9e-42;
Matches 198; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
>Y
1 ATGTTTCCAAAGGACCACTAACAGTCACCTTCTGGTATGGAGGAGCATGAGAGCTC 60
170 ATGTTTCTGAGGAGTGTATCGGCTGCTCACTCTGGGTTTCCAGGAGCATGAGAGCTC 239
61 TTTCATTGCTTTCTCTCTTCTGCTGCCATGTCAGCTCATTTGGAAGGCAAAAGGATAAT 120
230 TTTCATTGCTTTCTCTCTGCTGCTGCCAGGAGCTCATTTGGAAGGAGGAGCAAT 289
121 CAGTTCACTCTGGAACCAAGTCGCTGGGGAAGGTGTACAGGAGCTGTGTGTCGCCGAGGA 180
290 CAGTTCCTCTGGAACCAAGTCGCTGGGGAAGGTGTGTCAGGAGTGTGTCGCCGAGGGA 349
181 GTCCAGAGTGGGAGTGTGTGTTTCTGTTGACGGTGGGACAAAGTCACCTGTCTAAC 240

```

```

Db 350 GCCCAAGTCGTGTGTGTGTGTGTGTGTTCACATAGAGGGTGGACGAGTCCCATGTCCGAC 409
QY 241 TG 242
  ||
Db 410 TG 411

RESULT 11
LOCUS BF706397
DEFINITION 280725 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BF706397
VERSION BF706397.1 GI:1198058
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 456)
AUTHORS Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrrenkrug,S.C.,
Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A.,
Chitko-McKown,C.G., Pettea,G., Holt,I., Karamycheva,S., Liang,F.,
Quackenbush,J. and Keefe,J.W.
TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
JOURNAL Genome Res. 11 (4), 626-630 (2001)
MEDLINE 21180013
PUBMED 11282978
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smiththe@mail.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -mnscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCAGCG
Plate: 77 row: M column: 18
Seq primer: ATTAGTGACATATAG.
FEATURES
    source
        1..456
            /organism="Bos taurus"
            /mol_type="mRNA"
            /db_xref="taxon:9913"
            /tissue_type="pooled"
            /lab_host="DH10B"
            /clone_lib="MARC 3BOV"
            /note="Vector: pCMV SPORT6; Site 1: NotI, Site 2: SalI;
            Library made from pooled tissue from marrow, alveolar
            macrophage, ovary, fetal semitendinosus muscle, and fetal
            longissimus muscle."
ORIGIN
Query Match 28.3%; Score 119.8; DB 10; Length 456;
Best Local Similarity 91.4%; Pred No. 3.5e-26;
Matches 127; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 ATGTTTCCAAAGAGCAACTACAGTCACTTCTGCTGGGTATGAGGAGCATGAGGAAGTC 60
  |||||
Db 136 ATGTTTCCGAAGTAAATCTGGCTGTCACTGCTGGGTATGAGGAGCATGAGGAAGCTA 195
  |||||

QY 61 TTTCATTGCTTCTCTCTGTCGTCCCATCGACTATTTCGGAGGCAAAAGGATAAT 120
  |||||

Db 196 TTCTGTACTTCTCTCTGTCGTCCCATCGACTATTTCGGAGGCAAAAGGATAAT 255
  |||||

QY 121 CAGTTTCATCTGGAAACCCAG 139
  |||||
Db 256 CAGTTTCATCTGGAAACCCAG 274
  |||||

```

```

RESULT 12
LOCUS AK052119
DEFINITION
ACCESSION AK052119
VERSION AK052119.1 GI:26095006
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagao, K., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsu, T., Tashiro, H., Itoh, M.,
Suda, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujisaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
PANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6
AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurikawa, C., Matsuyama, T., Miyazaki, A., Murata, Y.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohnato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome.res@sc.riken.go.jp,

```

```

AK052119 557 bp mRNA linear HTC 20-SEP-2003
Mus musculus 12 days embryo eyeball cDNA, RIKEN full-length
enriched library, clone:D230048E01 product:weakly similar to
HYPOTHETICAL PROTEIN DJ110N13.1 IN CHROMOSOME 7 (FRAGMENT) [Homo
sapiens], full insert sequence.

```

```

AK052119
VERSION AK052119.1 GI:26095006
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```

REFERENCE 1
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagao, K., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsu, T., Tashiro, H., Itoh, M.,
Suda, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujisaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
PANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6
AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurikawa, C., Matsuyama, T., Miyazaki, A., Murata, Y.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohnato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome.res@sc.riken.go.jp,

```







Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CGAAGTGAAT. This library was created for the University of Iowa Brain Anatomy Project (BMAP); 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

RIGIN

```
Query Match      24.1%; Score 102.2; DB 14; Length 676;
Best Local Similarity 83.5%; Pred. No. 1.4e-20;
Matches 116; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

Y 1 ATGTTTCCAAAGACCACTAACAGTCACTTGTCTGGGTATGGAGGAGCATGAGGAGCTC 60
b 151 ATGTTTCTGAGGAGTGTCTGCTCTCACTCACTGGGTTCCAGGAGCATGAGGAGCTC 210

Y 61 TTTCATTGCTTTCTCTCTTGTGTCCATGCGCTCATTTTGGAGGCCAAAAGGATAAT 120
b 211 TTTCATTGCTTTCTCTCTTGTGTCCATGCGCTCATTTTGGAGGCCAAAAGGATAAT 120

Y 121 CAGTTCATCTGGAAACCG 139
b 271 CAGTTCCTCTGAAAACCG 289
```

search completed: February 23, 2004, 20:29:37  
Ob time : 973.64 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

DM nucleic - nucleic search, using sw model

Run on: February 23, 2004, 12:03:01 ; Search time 31.7428 Seconds  
(without alignments)  
7412.668 Million cell updates/sec

Title: US-10-022-710-5\_COPY\_1\_424

Perfect score: 424

Sequence: 1 atgtttccaaagagcaact.....geatggactgcagcaccgga 424

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents NA:\*
- 1: /cgn2\_6/ptodata/2/ina/5A COMB.seq.\*
  - 2: /cgn2\_6/ptodata/2/ina/5B COMB.seq.\*
  - 3: /cgn2\_6/ptodata/2/ina/6A COMB.seq.\*
  - 4: /cgn2\_6/ptodata/2/ina/6B COMB.seq.\*
  - 5: /cgn2\_6/ptodata/2/ina/PCFUS COMB.seq.\*
  - 6: /cgn2\_6/ptodata/2/ina/backfileseq1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	76	17.9	6378	4	US-09-620-312D-332
2	33.2	7.8	695	4	US-09-554-451-2
3	31	7.3	2866	4	US-09-724-864-30
4	30.8	7.3	1083	4	US-09-107-532A-869
5	30.6	7.2	239	4	US-09-642-703-33
6	30.6	7.2	288	4	US-09-642-703-34
7	30.6	7.2	1664976	4	US-08-916-421B-1
8	30.4	7.2	2824	4	US-07-757-022B-13
9	30.4	7.2	3066	4	US-07-757-022B-83
10	30.4	7.2	3117	4	US-07-757-022B-73
11	30.4	7.2	3148	4	US-07-757-022B-57
12	30.4	7.2	3420	4	US-07-757-022B-103
13	30.4	7.2	3813	4	US-07-757-022B-43
14	30.4	7.2	3936	4	US-07-757-022B-41
15	30.4	7.2	3942	4	US-07-757-022B-141
16	30.4	7.2	3945	4	US-07-757-022B-49
17	30.4	7.2	3963	4	US-07-757-022B-45
18	30.4	7.2	3963	4	US-07-757-022B-59
19	30.4	7.2	4065	4	US-07-757-022B-47
20	30.4	7.2	4086	4	US-07-757-022B-39
21	30.4	7.2	4092	4	US-07-757-022B-51
22	30.4	7.2	4215	4	US-07-757-022B-61
23	30.4	7.2	4847	4	US-10-164-595-57
24	30.4	7.2	5008	4	US-07-757-022B-1
25	30.4	7.2	5041	4	US-09-023-655-981
26	29.8	7.0	2342	4	US-09-620-312D-989
27	29.6	7.0	1022	3	US-09-189-035-4

Sequence 4, Appli  
Sequence 18, Appli  
Sequence 1349, Ap  
Sequence 14, Appli  
Sequence 1, Appli  
Sequence 7072, Ap  
Sequence 7027, Ap  
Sequence 5349, Ap  
Sequence 5192, Ap  
Sequence 6976, Ap  
Sequence 5283, Ap  
Sequence 327, App  
Sequence 14891, A  
Sequence 14740, A  
Sequence 15015, A  
Sequence 5, Appli  
Sequence 59, Appli  
Sequence 3, Appli

ALIGNMENTS

RESULT 1

US-09-620-312D-332  
; Sequence 332, Application US/09620312D  
; Patent No. 6569662  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Ren, Feiyan  
; APPLICANT: Chen, Rui-hong  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Yang, Yongchong  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Zhou, Ping  
; APPLICANT: Ma, Yunqing  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Wang, Zhiwei  
; APPLICANT: John Tillinghast  
; APPLICANT: Dmanac, Radoje T.  
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and  
; FILE REFERENCE: 784CIP2B  
; CURRENT APPLICATION NUMBER: US/09/620,312D  
; CURRENT FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 1105  
; SOFTWARE: pt FL\_genes Version 1.0  
; SEQ ID NO 332  
; LENGTH: 6378  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (271)..(5037)  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(6378)  
; OTHER INFORMATION: n = a,t,c or g  
US-09-620-312D-332

Query Match 17.9%; Score 76; DB 4; Length 6378;  
Best Local Similarity 59.8%; Pred. No. 3.7e-15;  
Matches 146; Conservative 0; Mismatches 95; Indels 3; Gaps 1;

QY 115 GATAATCAGTTTCATCTGGAAACACAGGTCCTGGGGAAGGTGTACAGGA---GACTGTGGT 171  
Db 229 GGTAATTTGGGAAGCTTCTGAGAAAGTCCATGGGCGGATGTATGGGAGATGAATGTGGT 288  
QY 172 CCGGAGAGTCCAGAGTCGGCGAGTGTGTGTTTTCATGTCACGGGTGGCAAGTCAC 231  
Db 289 CCGGAGGATCCAAACGAGGGCTGTGTGTGTGCTCATGTGGAGGATGGAATACACTG 348  
QY 232 CTGTCTAACTGTGTGAGAGCAACAGCGCTCCAAAGGAAGAAAGTTGTTTCCGAGTTGT 291  
Db 349 CATACTAACTGTAGCAGCGCGGAGACCCCAATACCAGCAGATTTGTTCAAGTTGC 408  
QY 292 GACTGGCAGAGTACCTCTTTCAGTGGAGGTTTCTGACTGGCAGCAGTGTGTGTTGT 351  
Db 409 GATTGGCACAAGAGTTGTACGACTGGAGACTGGACCTTGGAAATCAGTGTGACGGGTG 468  
QY 352 CCTT 355  
Db 469 ATTT 472

RESULT 2  
US-09-554-451-2/c  
; Sequence 2, Application US/09554451  
; Patent No. 6680207  
; GENERAL INFORMATION:  
; APPLICANT: Jonathan Paul MURPHY  
; TITLE OF INVENTION: Detection of Molecules in Samples  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pillsbury Winthrop, L.L.P.  
; STREET: 1100 New York Ave., N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: MS Word  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/GB98/03449  
; FILING DATE: No. 6680207ember 16, 1998  
; APPLICATION NUMBER: GB 9723955.2  
; FILING DATE: No. 6680207ember 14, 1997  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 695 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-554-451-2

Query Match 7.8%; Score 33.2; DB 4; Length 695;  
Best Local Similarity 53.0%; Pred. No. 0.29; Mismatches 63; Indels 0; Gaps 0;  
Matches 71; Conservative 0;  
QY 136 CCAGGTCCGTGGGAGGTGTACAGAGACTGTGTCGGGAGAGTCCAGAGTCGGCA 195  
Db 624 CCATGCTTTACGGAAGCAGTACAGACCGTAGTTTTTCAGCAGAGCGTCGTGTTGT 565  
QY 196 GTGTGGTGTTCATGTTCAGCGGTGGACAGTCACTGTCTTAATGTGGTGGAGCAAC 255  
Db 564 GAGAGTGTGTGCAATTTAGATAGTGTGTTTGAAGATCTGACCGGTACCGGAGAAC 505  
QY 256 AGGCCTCCAAAGA 269

Db 504 CGTCTTCCAGACGA 491  
RESULT 3  
US-09-724-864-30  
; Sequence 30, Application US/09724864  
; Patent No. 6380362  
; GENERAL INFORMATION:  
; APPLICANT: Watson, James D  
; APPLICANT: Murison, James G.  
; TITLE OF INVENTION: Polynucleotides, polypeptides expressed  
; TITLE OF INVENTION: by the polynucleotides and methods for their use.  
; FILE REFERENCE: 11000.105001  
; CURRENT APPLICATION NUMBER: US/09/724,864  
; CURRENT FILING DATE: 2000-11-28  
; PRIOR APPLICATION NUMBER: U.S. No. 6380362 60/171,678  
; PRIOR FILING DATE: 1999-12-23  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 30  
; LENGTH: 2866  
; TYPE: DNA  
; ORGANISM: Mouse  
US-09-724-864-30  
Query Match 7.3%; Score 31; DB 4; Length 2866;  
Best Local Similarity 50.3%; Pred. No. 3.3; Mismatches 75; Indels 0; Gaps 0;  
Matches 76; Conservative 0;  
QY 220 TGGACAAGTCACCTGTCTTAACCTGTGTGAGAGCAACAGCGCTCCAAAGAAAGTTGT 279  
Db 592 TGGACGAATCTCTTGGCCCTCGAATGTGAATGGAATGCAATGAAATGAAGGAACACAGG 651  
QY 280 TTCCGAGTTTGTGACTGGCAGACAGTCACTCTTTCAGTGGGAGTTTCTGACTGGCACCAC 339  
Db 652 GTCCAGCGCGGCACACACAGTTTACCAGTACCTCTCCATGGTACCTTCTCCGACGC 711  
QY 340 TGTGTGCTTGTTCCTTAGCTCGCGGTGAAG 370  
Db 712 AGACGGCAAGACATCACTCGCAGGTGGAG 742

RESULT 4  
US-09-107-532A-869  
; Sequence 869, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: PC  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,532A  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/085,598  
; FILING DATE: 14 May 1998  
; APPLICATION NUMBER: 60/051571  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:

NAME: Ariniello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 869:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1083 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (B) LOCATION 1...1083  
SEQUENCE DESCRIPTION: SEQ ID NO: 869:  
JS-09-107-532A-869

Query Match 7.3%; Score 30.8; DB 4; Length 1083;  
Best Local Similarity 57.1%; Pred. No. 2.3; Indels 0; Gaps 0;  
Matches 56; Conservative 0; Mismatches 42  
2y 48 CATGAGGAGCTCTTCTATTTGTTCTCTCTGTCGCCATGCGAGCTCATTTGGAAGG 107  
Db 504 CATGAATATCTTATTTTATTTGCTATTTTCTTAATTTCTTCAACATATAGGAATG 563  
2y 108 CAAAAGGATATCATGTTCTATCTGGAACACAGTCCGT 145  
Db 564 TCAAACTATTATGTCGTATATTGACTGCATTACCGT 601

RESULT 5  
US-09-642-703-33/c  
; Sequence 33, Application US/09642703  
; Patent No. 6524799  
; GENERAL INFORMATION:  
; APPLICANT: Walker, Michael G.  
; APPLICANT: Krasnow, Randi E.  
; TITLE OF INVENTION: SPARC-RELATED PROTEINS  
; FILE REFERENCE: PC-0015 CIP  
; CURRENT APPLICATION NUMBER: US/09/642,703  
; CURRENT FILING DATE: 2000-08-16  
; PRIOR APPLICATION NUMBER: 09/349,015  
; PRIOR FILING DATE: 1999-07-07  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PERL Program  
; SEQ ID NO 33  
; LENGTH: 239  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Incyte ID No. 6524799 701648524H1  
JS-09-642-703-33

Query Match 7.2%; Score 30.6; DB 4; Length 239;  
Best Local Similarity 50.3%; Pred. No. 1.3; Indels 0; Gaps 0;  
Matches 75; Conservative 0; Mismatches 74  
Qy 199 TGGTGTCTCATGTTGGACAGTCACTCTCTAACTGCTGGTGAGCAACAGG 258  
Db 154 TGGTGTCTTACAAATGGAAACAGAAACACACATAAGAGTTATTAAAGTGCAA 95  
Qy 259 CCTCCAAAGGAAGAAGTTGTTCCGAGTTTGACTGGCAGAGTCACTCTTTTCAGTGG 318  
Db 94 CATGGAGGACCACTTATGTATACATGGGCTGTGGTGGACCAAGGGCGGTGAAGT 35  
2y 319 GAGTTTCTGACTGGCCACTGTGTCT 347

Db 34 TAGGTGTCTCATGTCAGTCACTCTCTCTCT 6  
RESULT 6  
US-09-642-703-34/c  
; Sequence 34, Application US/09642703  
; Patent No. 6524799  
; GENERAL INFORMATION:  
; APPLICANT: Walker, Michael G.  
; APPLICANT: Krasnow, Randi E.  
; TITLE OF INVENTION: SPARC-RELATED PROTEINS  
; FILE REFERENCE: PC-0015 CIP  
; CURRENT APPLICATION NUMBER: US/09/642,703  
; CURRENT FILING DATE: 2000-08-16  
; PRIOR APPLICATION NUMBER: 09/349,015  
; PRIOR FILING DATE: 1999-07-07  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PERL Program  
; SEQ ID NO 34  
; LENGTH: 288  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Incyte ID No. 6524799 700306729H1  
US-09-642-703-34

Query Match 7.2%; Score 30.6; DB 4; Length 288;  
Best Local Similarity 50.3%; Pred. No. 1.4; Indels 0; Gaps 0;  
Matches 75; Conservative 0; Mismatches 74  
Qy 199 TGGTGTCTCATGTTGGACAGTCACTCTCTAACTGCTGGTGAGCAACAGG 258  
Db 227 TGGTGTCTTACAAATGGAAACAGAAACACACATAGAGTTATTAAAGTGCAA 168  
Qy 259 CCTCCAAAGGAAGAAGTTGTTCCGAGTTTGTGACTGGCAGTCACTCTTTTCAGTGG 318  
Db 167 CATGGAGGACCACTTATGTATCATGCTGGTGGTGGCCACGGCAGGCTGAAGT 108  
Qy 319 GAGTTTCTGACTGGCAGCAGTGTGTCT 347  
Db 107 TAGGTGTCTGATGGTCACTCTCTCTCT 79

RESULT 7  
US-08-916-421B-1  
; Sequence 1, Application US/08916421B  
; Patent No. 6503729  
; GENERAL INFORMATION:  
; APPLICANT: Bult et al.  
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus  
; Patent No. 6503729  
; TITLE OF INVENTION: jannaschii  
; FILE REFERENCE: PB275  
; CURRENT APPLICATION NUMBER: US/08/916,421B  
; CURRENT FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: US 60/024,428  
; PRIOR FILING DATE: 1996-08-22  
; NUMBER OF SEQ ID NOS: 3  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 1  
; LENGTH: 1664976  
; TYPE: DNA  
; ORGANISM: Methanococcus jannaschii  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (28222)..(28222)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc\_feature  
; LOCATION: (28257)..(28258)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc\_feature



```
, LOCATION: (84773)..(84773)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (622708)..(622708)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (657081)..(657081)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (657203)..(657203)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (674335)..(674335)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (682442)..(682442)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (713652)..(713652)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (741684)..(741684)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (779455)..(779455)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (779676)..(779676)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (855539)..(855539)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (871619)..(871619)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (1084830)..(1084830)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (1096846)..(1096846)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (1119881)..(1119881)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (1130881)..(1130881)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (1310988)..(1310988)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (1313224)..(1313224)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (1349473)..(1349473)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (1349491)..(1349491)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (1470091)..(1470091)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (1569020)..(1569020)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (1602912)..(1602912)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (1603734)..(1603734)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (1637998)..(1637998)
, OTHER INFORMATION: n equals a, t, c, or g
```

---

```
, LOCATION: (84773)..(84773)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (84808)..(84808)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (84812)..(84812)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (98120)..(98120)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (98159)..(98159)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (98239)..(98239)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (98266)..(98266)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (98343)..(98343)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (103998)..(103998)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (148948)..(148948)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (163385)..(163385)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (191989)..(191989)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (191995)..(191995)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (231980)..(231980)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (234187)..(234187)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (234220)..(234220)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (234814)..(234814)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (309398)..(309398)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (309418)..(309418)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (312837)..(312837)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (312993)..(312993)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (319226)..(319226)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (559167)..(559167)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (559241)..(559241)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (600992)..(600992)
```

```

; NAME/KEY: misc feature
; LOCATION: (1664854)..(1664855)
; OTHER INFORMATION: n equals a, t, c, or g
JS-08-916-421B-1

Query Match      7.2%; Score 30.6; DB 4; Length 1664976;
Best Local Similarity 65.2%; Pred. No. 94;
Matches 45; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

2Y 99 TTGAGGCAAAAGATATACGTTTCATCTGGAACACAGGTCGGTGGGGAAGGTGTAC 158
Db 1031419 TTGGAATCCAGAAAGATAGTATTTATTTGACCAAGTTCCTGCTGATAGTATAA 1031478
2Y 159 AGGAGACTG 167
Db 1031479 AAGCGGCTG 1031487

RESULT 8
US-07-757-022B-13/c
; Sequence 13, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELEPHONE: (617) 876-1170
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2824 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS

```





TELEFAX: (617)876-5851  
INFORMATION FOR SEQ ID NO: 43:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3813 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..3810  
US-07-757-022B-43

Query Match 7.2%; Score 30.4; DB 4; Length 3813;  
Best Local Similarity 55.7%; Pred. No. 6;  
Matches 78; Conservative 0; Mismatches 61; Indels 1; Gaps 1;  
QY 58 CTCCTTCTATGCTTCTCTCTGTCGCCATGCGAGCTCATTTGGAGGCAAAAGGAT 117  
Db 2365 CTCCTTCTATGCTTCTCTCTGTCGCCATGCGAGCTCATTTGGAGGCAAAAGGAT 117  
QY 118 ATCAGTTCACTGGAACCCAGGTCGGTGGGAAGGTGTACAGGAGCTGTGTCGCCGA 177  
Db 2305 TCTTAGTT-GTAGGTACACAGGTCCTTGGGACTGTTTCAAGAGCTTTTGGTGGGT 2247  
QY 178 GGAGTCCAGAGTCGGGCAGT 197  
Db 2246 TCTGCAGAAAGCTCAGGAGT 2227

RESULT 14  
US-07-757-022B-41/c  
Sequence 41, Application US/07757022B  
Patent No. 6433142  
GENERAL INFORMATION:  
APPLICANT: Gesner, Thomas G.  
APPLICANT: Clark, Stephen C.  
APPLICANT: Turner, Katherine  
APPLICANT: Hewick, Rodney M.  
TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
NUMBER OF SEQUENCES: 143  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/757,022B  
FILING DATE: 19910910  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/643,502  
FILING DATE: 18-JAN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/546,114  
FILING DATE: 29-JUN-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/457,196  
FILING DATE: 29-DEC-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/390,901  
FILING DATE: 08-AUG-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Cseri, Luann  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: GI 5190

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)876-1170  
TELEFAX: (617)876-5851  
INFORMATION FOR SEQ ID NO: 41:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3936 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..3933  
US-07-757-022B-41

Query Match 7.2%; Score 30.4; DB 4; Length 3936;  
Best Local Similarity 55.7%; Pred. No. 6.1;  
Matches 78; Conservative 0; Mismatches 61; Indels 1; Gaps 1;  
QY 58 CTCCTTCTATGCTTCTCTCTGTCGCCATGCGAGCTCATTTGGAGGCAAAAGGAT 117  
Db 2488 CTCCTTCTATGCTTCTCTCTGTCGCCATGCGAGCTCATTTGGAGGCAAAAGGAT 117  
QY 118 ATCAGTTCACTGGAACCCAGGTCGGTGGGAAGGTGTACAGGAGCTGTGTCGCCGA 177  
Db 2428 TCTTAGTT-GTAGGTACACAGGTCCTTGGGACTGTTTCAAGAGCTTTTGGTGGGT 2370  
QY 178 GGAGTCCAGAGTCGGGCAGT 197  
Db 2369 TCTGCAGAAAGCTCAGGAGT 2350

RESULT 15  
US-07-757-022B-141/c  
Sequence 141, Application US/07757022B  
Patent No. 6433142  
GENERAL INFORMATION:  
APPLICANT: Gesner, Thomas G.  
APPLICANT: Clark, Stephen C.  
APPLICANT: Turner, Katherine  
APPLICANT: Hewick, Rodney M.  
TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
NUMBER OF SEQUENCES: 143  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/757,022B  
FILING DATE: 19910910  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/643,502  
FILING DATE: 18-JAN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/546,114  
FILING DATE: 29-JUN-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/457,196  
FILING DATE: 29-DEC-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/390,901  
FILING DATE: 08-AUG-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Cseri, Luann

```

/ REGISTRATION NUMBER: 31,822
/ REFERENCE/DOCKET NUMBER: GI 5190
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (617)876-1170
/ TELEFAX: (617)876-5851
/ INFORMATION FOR SEQ ID NO: 141:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 3942 base pairs
/ TYPE: NUCLEIC ACID
/ STRANDEDNESS: double
/ TOPOLOGY: unknown
/ MOLECULE TYPE: cDNA
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 1..3939
/ US-07-757-022B-141

Query Match      7.2%; Score 30.4; DB 4; Length 3942;
Best Local Similarity 55.7%; Pred.No.6.1;
Matches 78; Conservative 0; Mismatches 61; Indels 1; Gaps 1;

Qy 58 CTCTTTCTATGCTTTCTCTTCTGCTGCCATGCCAGCTCATTTGGAAGGCCAAAAAGGAT 117
Db 2494 CTCTTTCTGTTGTTCTTCTCTTTAGTCTGTTAGTCATTTCAGGTTTAGTCGCTGCAGGAG 2435

Qy 118 AATCAGTTCTCTGGAACAGGTCCTGGGGAAGGTGTACRGGAGACTGTGTCCTCCGGA 177
Db 2434 TCTTAGTT-GTAGGTACACCAAGTTCCTTGGGACTGTTTCAAGAGCTTTTGGTGTGGGT 2376

Qy 178 GGAGTCCAGAGTCGGGCAGT 197
Db 2375 TCTGCAGAAAGCTCAGGAGT 2356
```

Search completed: February 23, 2004, 20:35:49  
Job time : 35.7428 secs



GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

XM protein - protein search, using sw model

Run on: February 23, 2004, 20:29:46 ; Search time 12.9894 Seconds  
(without alignments)  
3088.802 Million cell updates/sec

Title: US-10-022-710-6\_COPY\_1\_142

Perfect score: 830

Sequence: 1 MFPSKNTLTCWWSMRKL.....EVKPTAEQVTAQHGLOHRM 142

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- A\_Geneseq\_29Jan04:\*
- 1: Geneseqp1980s:\*
  - 2: Geneseqp1990s:\*
  - 3: Geneseqp2000s:\*
  - 4: Geneseqp2001s:\*
  - 5: Geneseqp2002s:\*
  - 6: Geneseqp2003as:\*
  - 7: Geneseqp2003bs:\*
  - 8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	580.5	69.9	1574	7 ADE08053	Ade08053 Novel pro
2	573	69.0	247	4 AAM70792	Aam70792 Human bon
3	573	69.0	247	4 AAM58317	Aam58317 Human bra
4	327.5	39.5	933	7 ADB64774	Adb64774 Human pro
5	297.5	35.8	1624	5 AAE23979	Aae23979 Human Lp2
6	289	34.8	277	4 AAM28984	Aam28984 Peptide #
7	289	34.8	277	4 ABG50341	Abg50341 Human liv
8	289	34.8	617	3 AAB42496	Aab42496 Human ORF
9	253.5	30.5	1549	6 ABR58455	Abr58455 Human NOV
10	252.5	30.4	1490	6 ABR58461	Abr58461 Human NOV
11	252.5	30.4	1490	6 ABR58454	Abr58454 Human NOV
12	252.5	30.4	1545	6 ABR58453	Abr58453 Human NOV
13	252.5	30.4	1588	4 AAB20135	Aab20135 Secreted
14	252.5	30.4	1588	4 AAM39295	Aam39295 Human pol
15	252.5	30.4	1588	6 ABR58450	Abr58450 Human NOV
16	252.5	30.4	1588	6 ABR58456	Abr58456 Human NOV
17	252.5	30.4	1588	6 ABR58459	Abr58459 Human NOV
18	252.5	30.4	1588	6 ABR58462	Abr58462 Human NOV
19	128.5	15.5	1057	6 ABP71970	Abp71970 Human agg
20	128.5	15.5	1095	6 ABR40094	Abr40094 Human ADA
21	128.5	15.5	1122	6 ABP71971	Abp71971 Human agg
22	128.5	15.5	1145	6 ABP71974	Abp71974 Human agg
23	127	15.3	622	4 AAB5346	Aab5346 Human pro
24	127	15.3	1363	7 ADD44592	Add44592 Polypepti
25	127	15.3	1505	5 AAU72897	Aau72897 Human met

26	127	15.3	1602	5 ABG30702	Abg30702 Human agg
27	127	15.3	1629	5 ABG30703	Abg30703 Human agg
28	127	15.3	1629	5 AAO14448	Aao14448 Human ADA
29	127	15.3	1907	5 AAU77133	Aau77133 Human pro
30	127	15.3	1916	5 AAE19173	Aae19173 Human pro
31	127	15.3	1934	4 AAB72301	Aab72301 Human ADA
32	127	15.3	1935	5 AAU72896	Aau72896 Human met
33	124	14.9	513	6 ABU54545	Abu54545 Human NOV
34	122	14.7	1882	4 AAB72286	Aab72286 Human ADA
35	121.5	14.6	226	4 AAE10824	Aae10824 Human gen
36	121.5	14.6	465	5 AAE21004	Aae21004 Human pro
37	121.5	14.6	476	5 AAE21001	Aae21001 Human pro
38	121.5	14.6	807	4 ABG14803	Abg14803 Novel hum
39	121.5	14.6	969	5 AAE21003	Aae21003 Human pro
40	121.5	14.6	980	5 AAE21000	Aae21000 Human pro
41	121.5	14.6	1035	5 ABB98128	Abb98128 Human PMM
42	121.5	14.6	1072	6 ABU12107	Abu12107 Human pro
43	121.5	14.6	1094	5 AAU72900	Aau72900 Human met
44	121.5	14.6	1213	5 AAE21002	Aae21002 Human pro
45	121.5	14.6	1224	5 AAE20999	Aae20999 Human pro

ALIGNMENTS

RESULT 1  
ADE08053  
ID ADE08053 standard; protein; 1574 AA.  
XX

AC ADE08053;

XX 29-JAN-2004 (first entry)

XX Novel protein (useful for identifying genetic disorders) #208.

XX novel gene; novel protein; tissue marker; molecular weight marker;

XX chromosome marker; genetic disorder.

XX Unidentified.

XX WO2003054152-A2.

XX 03-JUL-2003.

XX 10-DEC-2002; 2002WO-US039555.

XX 10-DEC-2001; 2001US-0339739P.

XX 11-DEC-2001; 2001US-0339453P.

XX 14-MAR-2002; 2002US-0365091P.

XX 12-APR-2002; 2002US-0372381P.

XX 22-APR-2002; 2002US-0372615P.

XX 24-APR-2002; 2002US-0376045P.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;

XX Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang Z;

XX Ma Y, Wang D, Chen R, Xu C, Boyle BJ;

XX WPI; 2003-569235/53.

XX N-PSDB; ADE07142.

XX New polynucleotides, useful for expressing recombinant proteins for

XX analysis, characterization or therapeutic use, or as markers for tissues

XX in which the corresponding protein is preferentially expressed.

XX Claim 20; SEQ ID NO 1119; 1177pp; English.

XX The invention comprises the amino acid and coding sequences of novel

XX proteins. The DNA and protein sequences of the invention are useful as:

XX markers for tissues in which the corresponding protein is preferentially

CC expressed, as molecular weight markers on gels, as chromosome markers or  
CC tags; to identify chromosomes or to map related gene positions; and to  
CC compare with endogenous DNA sequences in patients to identify potential  
CC genetic disorders. The present amino acid sequence represents a protein  
CC of the invention.  
XX  
SQ Sequence 1574 AA;

Query Match 69.9%; Score 580.5; DB 7; Length 1574;  
Best Local Similarity 99.0%; Pred. No. 2.7e-50;  
Matches 98; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
QY 45 KP-GPWGCTGCGGGVQSVRAVWCFHVDGWTSHLSNCGESNRPKRSFCRVCDWHSDDL 103  
DB 10 KPHGPGWGRCTGCGGGVQSVRAVWCFHVDGWTSHLSNCGESNRPKRSFCRVCDWHSDDL 69  
QY 104 FQWEVSDWHHCVLVPYARGEVKPRTAEVCVTAQHGLOHRM 142  
DB 70 FQWEVSDWHHCVLVPYARGEVKPRTAEVCVTAQHGLOHRM 108

RESULT 2  
AAM70792  
ID AAM70792 standard; protein; 247 AA.  
XX  
AC AAM70792;  
DT 06-NOV-2001 (first entry)  
DE Human bone marrow expressed probe encoded protein SEQ ID NO: 31098.  
KW Human; bone marrow expressed exon; gene expression analysis; probe;  
KW microarray; cancer; leukaemia; lymphoma; myeloma.  
XX  
OS Homo sapiens.  
XX WO200157276-A2.  
XX 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US000668.  
XX  
XX 04-FEB-2000; 2000US-0180312P.  
XX 26-MAY-2000; 2000US-0207456P.  
XX 30-JUN-2000; 2000US-00608408.  
XX 03-AUG-2000; 2000US-00632366.  
XX 21-SEP-2000; 2000US-0234687P.  
XX 27-SEP-2000; 2000US-0236359P.  
XX 04-OCT-2000; 2000GB-00024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488900/53.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for analyzing  
XX gene expression in human bone marrow.  
XX  
XX Example 4; SEQ ID NO 31098; 658pp + Sequence Listing; English.

Query Match 69.0%; Score 573; DB 4; Length 247;  
Best Local Similarity 100.0%; Pred. No. 2e-50;  
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 48 PWGRTGDCGPGGVQSVRAVWCFHVDGWTSHLSNCGESNRPKRSFCRVCDWHSDDL 107  
DB 1 PWGRTGDCGPGGVQSVRAVWCFHVDGWTSHLSNCGESNRPKRSFCRVCDWHSDDL 60  
QY 108 VSDWHHCVLVPYARGEVKPRTAEVCVTAQHGLOHRM 142  
DB 61 VSDWHHCVLVPYARGEVKPRTAEVCVTAQHGLOHRM 95

Query Match 69.0%; Score 573; DB 4; Length 247;  
Best Local Similarity 100.0%; Pred. No. 2e-50;  
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 48 PWGRTGDCGPGGVQSVRAVWCFHVDGWTSHLSNCGESNRPKRSFCRVCDWHSDDL 107  
DB 1 PWGRTGDCGPGGVQSVRAVWCFHVDGWTSHLSNCGESNRPKRSFCRVCDWHSDDL 60  
QY 108 VSDWHHCVLVPYARGEVKPRTAEVCVTAQHGLOHRM 142  
DB 61 VSDWHHCVLVPYARGEVKPRTAEVCVTAQHGLOHRM 95

QY 48 PWGRTGDCGPGGVQSVRAVWCFHVDGWTSHLSNCGESNRPKRSFCRVCDWHSDDL 107  
DB 1 PWGRTGDCGPGGVQSVRAVWCFHVDGWTSHLSNCGESNRPKRSFCRVCDWHSDDL 60  
QY 108 VSDWHHCVLVPYARGEVKPRTAEVCVTAQHGLOHRM 142  
DB 61 VSDWHHCVLVPYARGEVKPRTAEVCVTAQHGLOHRM 95

RESULT 3  
AAM58317  
ID AAM58317 standard; protein; 247 AA.  
XX  
AC AAM58317;  
DT 05-NOV-2001 (first entry)  
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 30422.  
KW Human; brain expressed exon; gene expression analysis; probe; microarray;  
KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.  
XX  
OS Homo sapiens.  
XX WO200157275-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US000667.  
XX  
XX 04-FEB-2000; 2000US-0180312P.  
XX 26-MAY-2000; 2000US-0207456P.  
XX 30-JUN-2000; 2000US-00608408.  
XX 03-AUG-2000; 2000US-00632366.  
XX 21-SEP-2000; 2000US-0234687P.  
XX 27-SEP-2000; 2000US-0236359P.  
XX 04-OCT-2000; 2000GB-00024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-483446/52.  
XX  
XX Single exon nucleic acid probes for analyzing gene expression in human  
XX brains.  
XX  
XX Example 4; SEQ ID NO 30422; 650pp + Sequence Listing; English.

The present invention provides a number of single exon nucleic acid  
probes which are derived from genomic sequences expressed in the human  
brain. They can be used to measure gene expression in brain cell samples,  
which may enable the diagnosis and improved treatment of nervous system  
diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia, of  
epilepsy and cancers. The present sequence is a protein encoded by one of  
the probes of the invention  
XX  
XX Sequence 247 AA;  
Query Match 69.0%; Score 573; DB 4; Length 247;  
Best Local Similarity 100.0%; Pred. No. 2e-50;  
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 48 PWGRTGDCGPGGVQSVRAVWCFHVDGWTSHLSNCGESNRPKRSFCRVCDWHSDDL 107  
DB 1 PWGRTGDCGPGGVQSVRAVWCFHVDGWTSHLSNCGESNRPKRSFCRVCDWHSDDL 60  
QY 108 VSDWHHCVLVPYARGEVKPRTAEVCVTAQHGLOHRM 142  
DB 61 VSDWHHCVLVPYARGEVKPRTAEVCVTAQHGLOHRM 95

Query Match 69.0%; Score 573; DB 4; Length 247;  
Best Local Similarity 100.0%; Pred. No. 2e-50;  
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 48 PWGRTGDCGPGGVQSVRAVWCFHVDGWTSHLSNCGESNRPKRSFCRVCDWHSDDL 107  
DB 1 PWGRTGDCGPGGVQSVRAVWCFHVDGWTSHLSNCGESNRPKRSFCRVCDWHSDDL 60  
QY 108 VSDWHHCVLVPYARGEVKPRTAEVCVTAQHGLOHRM 142  
DB 61 VSDWHHCVLVPYARGEVKPRTAEVCVTAQHGLOHRM 95

Query Match 69.0%; Score 573; DB 4; Length 247;  
Best Local Similarity 100.0%; Pred. No. 2e-50;  
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 48 PWGRTGDCGPGGVQSVRAVWCFHVDGWTSHLSNCGESNRPKRSFCRVCDWHSDDL 107  
DB 1 PWGRTGDCGPGGVQSVRAVWCFHVDGWTSHLSNCGESNRPKRSFCRVCDWHSDDL 60  
QY 108 VSDWHHCVLVPYARGEVKPRTAEVCVTAQHGLOHRM 142  
DB 61 VSDWHHCVLVPYARGEVKPRTAEVCVTAQHGLOHRM 95

QY	22	LLLSLLL-----SHAHLBEGKDNQFIWKDPPGRCTGD--CGPGGVQSRVAVWCVFHDVGH	75
DB	34	LLLLLLLLPPGAGRAAAGGEAAEAPTLILWKTPGPGRCMGDECGPGGGIQTTRAVVMAHVEGVT	93
QY	76	SHLNSCGESNRPKRCFCFVCDWHSDFOWEVSDDWHHCVLVPAARGEVKPRTAECVTAQ	135
DB	94	TLHTNCKAERFNNQNCFKVCDWHKELYDRLGPNRCOPV-ISKSLKP--LECIKE	150
QY	136	HGLQHR 141	
DB	151	EGIQVR 156	
RESULT 5			
AAE23979	AAE23979 standard; protein; 1624 AA.		
XX	AC	AAE23979;	
XX	DT	23-SEP-2002 (first entry)	
XX	DE	Human LP217 secreted protein.	
XX	XX	Human; secreted protein; atherosclerosis; Alzheimer's disease; LP217;	
XX	KW	diabetic retinopathy; severe combined immunodeficiency; pancreatitis;	
XX	KW	rheumatoid arthritis; colorectal adenoma; haemolytic anaemia; cancer;	
XX	KW	reperfusion injury; arteriosclerosis; wound healing; transgenic animal;	
XX	KW	gene therapy; neoplasm; transgenic; psoriasis; ischaemia; carcinoma;	
XX	KW	chromosome 7p21-p22.	
XX	OS	Homo sapiens.	
XX	XX	Key	Location/Qualifiers
XX	FF	Peptide	1. .15
FT	FT		/label= Signal_peptide
FT	FT	Protein	16..1624
FT	FT		/note= "Mature human LP217 secreted protein"
XX	XX	WO200226801-A2.	
XX	PN		
XX	PD	04-APR-2002.	
XX	PF	14-SEP-2001; 2001WO-US026026.	
XX	PR	28-SEP-2000; 2000US-0236088P.	
XX	PA	(ELIL ) LILLY & CO ELI.	
XX	PI	Su EW, Wang H;	
XX	XX	WPI; 2002-471259/50.	
DR	DR	N-PSDB; AAD38694.	
XX	XX	Novel proteins and polynucleotides of secreted proteins useful for	
XX	PT	treating various diseases e.g. rheumatoid arthritis, cancer, psoriasis,	
XX	PT	diabetic retinopathy, arteriosclerosis, ischemia or reperfusion injury.	
XX	PS	Claim 9; Page 117-123; 145pp; English.	
XX	XX	The invention relates to human secreted polypeptides designated LP095,	
CC	CC	LP191, LP217, LP220, LP221, LP222, LP229, LP237 or LP238 and nucleic acid	
CC	CC	molecules encoding such polypeptides. Novel secreted proteins of the	
CC	CC	invention are used for treating diseases such as atherosclerosis,	
CC	CC	Alzheimer's disease, diabetic retinopathy, psoriasis, pancreatitis,	
CC	CC	arteriosclerosis, rheumatoid arthritis, colorectal adenomas, severe	
CC	CC	combined immunodeficiency, ischaemia, carcinomas, haemolytic anaemia,	
CC	CC	reperfusion injury, neoplasms and cancer especially liver cancer. They	
CC	CC	are also used for wound healing. Polynucleotides of the invention can be	
CC	CC	used to generate transgenic animals or knock out animals, which in turn,	
CC	CC	are useful in the development and screening of therapeutically useful	
CC	CC	reagents for use in the treatment of diseases associated with LP	
CC	CC	polypeptide associated activity. They are also used in gene therapy. The	
CC	CC	present sequence is human LP217 secreted protein. LP217 gen is located	

```
CC on chromosome 7p21-p22
XX Sequence 1624 AA;
SQ Query Match 35.8%; Score 297.5; DB 5; Length 1624;
Best Local Similarity 38.8%; Pred. No. 2.6e-21;
Matches 52; Conservative 22; Mismatches 27; Indels 33; Gaps 5;

QY 9 VTCWVRSMRKLFLLSLLSHAAHLEKQDNQFIWKPGPWGRCTGD-CGPGGVQSRVAVW 67
DB 22 VPCNI-----FIL-----TGPNRCMGDECGPGGIQTRAVW 52

QY 68 CFHYDGTWTHSLNCGESNRPPKERSCFRVCMDHSDLFQNEVSDMHCHVLPYARGEVKPR 127
DB 53 CAHVEGTTTLTNCQAERPNQCNCFKVCVDHKLKELYDRLGPNWQCQPV-ISKSLKRP-110

QY 128 TAEVCVTAQHGLQHR 141
DB 111 -LECIKGEGIQVR 123

RESULT 6
AAM28984
ID AAM28984 standard; protein; 277 AA.
XX
AC AAM28984;
XX
XX 17-OCT-2001 (first entry)
XX
DE Peptide #3021 encoded by probe for measuring placental gene expression.
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW Genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000663.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX
XX 26-MAY-2000; 2000US-0207456P.
XX
XX 30-JUN-2000; 2000US-0060840P.
XX
XX 03-AUG-2000; 2000US-00632366.
XX
XX 21-SEP-2000; 2000US-0234687P.
XX
XX 27-SEP-2000; 2000US-0236359P.
XX
XX 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488997/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human placenta.
XX
XX Claim 27; SEQ ID NO 29253; 654pp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENPs)
XX see AAI31315-AI57546). The present sequence is a peptide encoded by one
XX such probe. The probes are useful for producing a microarray for
XX predicting, measuring and displaying gene expression in samples derived
XX from human placenta. The probes are useful for antenatal diagnosis of
XX human genetic disorders
XX
XX Sequence 277 AA;
XX
XX Query Match 34.8%; Score 289; DB 4; Length 277;
XX Best Local Similarity 48.4%; Pred. No. 2.8e-21;
XX Matches 46; Conservative 20; Mismatches 25; Indels 4; Gaps 3;

QY 48 PWGRCCTGD-CGPGGVQSRVAVWCFHYDGTWTHSLNCGESNRPPKERSCFRVCMDHSDLFQW 106
DB 1 PWGRCMGDECGPGGIQTRAVWCAHVEGTTTLTNCQAERPNQCNCFKVCVDHKLKELYDWM 60

QY 107 EYSDMHCHVLPYARGEVKPRTAECVTAQHGLQHR 141
DB 61 RLGPWNQCQPV-ISKSLKRP--LECIKGEGIQVR 92

RESULT 7
ABG50341
ID ABG50341 standard; peptide; 277 AA.
XX
AC ABG50341;
XX
XX 25-FEB-2003 (first entry)
XX
XX Human liver peptide, SEQ ID No 28989.
XX
XX Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
XX hypercholesterolaemia; coronary heart disease.
XX
XX Homo sapiens.
XX
XX WO200157273-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000664.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX
XX 26-MAY-2000; 2000US-0207456P.
XX
XX 30-JUN-2000; 2000US-0060840P.
XX
XX 03-AUG-2000; 2000US-00632366.
XX
XX 21-SEP-2000; 2000US-0234687P.
XX
XX 27-SEP-2000; 2000US-0236359P.
XX
XX 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488998/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human adult liver.
XX
XX Claim 27; SEQ ID NO 28989; 658pp; English.
XX
XX The invention relates to a single exon nucleic acid probe (SENP) (I) for
XX measuring human gene expression in a sample derived from human adult
XX liver, comprising one of 13109 defined nucleotide sequences given in the
XX specification (or complements/ fragments). The probe hybridises at high
XX stringency to a nucleic acid molecule expressed in the human adult liver.
XX (I) may be used for predicting, measuring and displaying gene expression
XX in samples derived from human adult liver. The genes identified may be
XX involved in genetic liver diseases such as cirrhosis,
XX hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
XX associated with coronary heart disease. ABG47348-ABG59930 represent human
XX liver single exon encoded peptides of the invention. Note: The sequence
XX information for this patent does not appear in the printed specification
XX but was obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 277 AA;
XX
XX Query Match 34.8%; Score 289; DB 4; Length 277;
XX Best Local Similarity 48.4%; Pred. No. 2.8e-21;
XX Matches 46; Conservative 20; Mismatches 25; Indels 4; Gaps 3;

QY 48 PWGRCCTGD-CGPGGVQSRVAVWCFHYDGTWTHSLNCGESNRPPKERSCFRVCMDHSDLFQW 106
DB 1 PWGRCMGDECGPGGIQTRAVWCAHVEGTTTLTNCQAERPNQCNCFKVCVDHKLKELYDWM 60
```

CC		storage, systemic lupus erythematosus, severe combined immunodeficiency (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to enhance coagulation; to inhibit thrombosis; and as a contraceptive
XX		
SQ		Sequence 617 AA;
	Query Match	34.8%; Score 289; DB 3; Length 617;
	Best Local Similarity	48.4%; Pred. No. 6.7e-21;
	Matches 46; Conservative 20; Mismatches 25; Indels 4; Gaps 3	
Qy	48	PWGRCTGD-CGPGVGQSRAYVCFHVDGTTSHLNCGESNEPPKERSCFRVCDSFLFOW 106
Dd	1	PWRCNCGDECPGGIQTAVWCVAHVGEWTILHTNCQARPNNQCNCFKVCDSWHKELYDW 60
Qy	107	EVSDDHHCHVLVPYARGEVKPRTAECVTAOHLQHR 141
Dd	61	RLGPNWQQP-V-ISKLEKP--LECIGKEGIQVR 92
	RESULT 9	
ID	ABRS8455	
IR	ABRS8455 standard; protein; 1549 AA.	
XX	AC AC	ABRS8455;
XX	XX XX	
DT	07-JUL-2003 (first entry)	
XX	XX XX	Human NOV47f.
DE	XX XX	Human; NOV; antidiabetic; anorectic; antibacterial; virucide; immunomodulator; cytostatic; nootropic; neuroprotective; dyslipidaemia; antiparkinsonian; antilipaeamic; gene therapy; metabolic disorder; diabetes; obesity; infection; cachexia; cancer; Parkinson's disease; neurodegenerative disorder; Alzheimer's disease; immune disorder; haematopoietic disorder.
KW	KW KW	Homo sapiens.
OS	OS OS	WO2003029423-A2.
FN	FN FN	10-APR-2003.
PD	PD PD	02-OCT-2002; 2002WC-US031358.
XX	XX XX	02-OCT-2001; 2001US-0326483P.
XX	XX XX	05-OCT-2001; 2001US-0327342P.
PR	PR PR	09-OCT-2001; 2001US-0327917P.
PR	PR PR	09-OCT-2001; 2001US-0328029P.
PR	PR PR	09-OCT-2001; 2001US-0328044P.
PR	PR PR	09-OCT-2001; 2001US-0328056P.
PR	PR PR	12-OCT-2001; 2001US-0328849P.
PR	PR PR	15-OCT-2001; 2001US-032914P.
PR	PR PR	17-OCT-2001; 2001US-0330142P.
PR	PR PR	22-OCT-2001; 2001US-0334058P.
PR	PR PR	24-OCT-2001; 2001US-0339266P.
PR	PR PR	24-OCT-2001; 2001US-0343259P.
PR	PR PR	29-OCT-2001; 2001US-0349575P.
PR	PR PR	01-NOV-2001; 2001US-0346357P.
PR	PR PR	12-APR-2002; 2002US-0371972P.
PR	PR PR	12-APR-2002; 2002US-0371980P.
PR	PR PR	17-APR-2002; 2002US-0373261P.
PR	PR PR	19-APR-2002; 2002US-0373805P.
PR	PR PR	23-APR-2002; 2002US-0374738P.
PR	PR PR	16-MAY-2002; 2002US-0381101P.
PR	PR PR	17-MAY-2002; 2002US-0381635P.
PR	PR PR	29-MAY-2002; 2002US-0383830P.
PR	PR PR	01-OCT-2002; 2002US-0026283P.
XX	XX XX	(CURA-) CURAGEN CORP.
PA	PA PA	Aalsbroek JP, Anderson DW, Boldog FL, Burgess CE, Catterton E;
PI	PI PI	

PI Edinger SR, Ellerman K, Gerlach VL, Gorman L, Guo X, Ji W;  
 PI Kekuda R, Leach MD, Li L, Miller CE, Patturajan M, Rieger DK;  
 PI Rothenberg ME, Shimkets RA, Smithson G, Spytek KA, Taupier RJ;  
 PI Vernet CAM, Voss EZ, Zerhusen BD, Zhong M;  
 XX WPI: 2003-381625/36.  
 DR N-PSDB; ACC72167.  
 DR  
 XX  
 XX NOVX polypeptides and nucleic acids useful for diagnosing, preventing or  
 PT treating NOVX-associated disorders, e.g. diabetes, obesity, cancer or  
 PT dyslipidemia, and in chromosome mapping, tissue typing or  
 PT pharmacogenomics.  
 XX  
 XX Claim 1; Page 260; 487pp; English.  
 XX  
 XX The present invention relates to novel human NOV proteins and their  
 CC coding sequences (ACC72075-ACC72181 and ABR58363-ABR58469). The NOV  
 CC proteins are useful in manufacturing a medicament for treating a syndrome  
 CC associated with a human disease. The NOV proteins and coding sequences  
 CC may be used to diagnose, treat or prevent metabolic disorders such as  
 CC diabetes or obesity, infections, cachexia, cancer, neurodegenerative  
 CC disorders such as Alzheimer's disease or Parkinson's disease, immune  
 CC disorders, haematopoietic disorders and various dyslipidaemias  
 XX  
 XX Sequence 1549 AA;  
 SQ  
 Query Match 30.5%; Score 253.5; DB 6; Length 1549;  
 Best Local Similarity 43.5%; Pred. No. 8e-17; Mismatches 28; Indels 3; Gaps 2;  
 Matches 40; Conservative 21; Mismatches 28; Indels 3; Gaps 2;  
 QY 50 GRCTGDCGGVQSVRAVWCFHVDGWTSHLSNCGESNRPKERSCFRVCDHSDLFQWEVS 109  
 Db 1 GTMGDECGGGGIGTAVVCAHVEGWTTLHTNCKQAEPPNQQNCFKVDWHKELYDRLG 60  
 QY 110 DWHCULVPYARGEVKPTAEVCVTAHQGLQHR 141  
 Db 61 PNNOCQPV-ISKLEKP--LECIRGEEGIQVR 89  
 RESULT 10  
 ABR58461  
 ID ABR58461 standard; protein; 307 AA.  
 AC ABR58461;  
 XX  
 XX 07-JUL-2003 (first entry)  
 DT Human NOV47L.  
 DE  
 XX Human; NOV; antidiabetic; anorectic; antibacterial; virucide;  
 KW immunomodulator; cytostatic; nootropic; neuroprotective; dyslipidaemia;  
 KW antiparkinsonian; antilipaeamic; gene therapy; metabolic disorder;  
 KW diabetes; obesity; infection; cachexia; cancer; Parkinson's disease;  
 KW neurodegenerative disorder; Alzheimer's disease; immune disorder;  
 KW haematopoietic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO2003029423-A2.  
 XX  
 PD 10-APR-2003.  
 XX  
 FF 02-OCT-2002; 2002WO-US031358.  
 XX  
 XX 02-OCT-2001; 2001US-0326483P.  
 PR 05-OCT-2001; 2001US-0327342P.  
 PR 09-OCT-2001; 2001US-0327917P.  
 PR 09-OCT-2001; 2001US-0328029P.  
 PR 09-OCT-2001; 2001US-0328044P.  
 PR 09-OCT-2001; 2001US-0328056P.  
 PR 12-OCT-2001; 2001US-0328849P.  
 PR 15-OCT-2001; 2001US-0329414P.  
 PR 17-OCT-2001; 2001US-0330142P.

PR 22-OCT-2001; 2001US-0341058P.  
 PR 24-OCT-2001; 2001US-0339266P.  
 PR 24-OCT-2001; 2001US-0343629P.  
 PR 29-OCT-2001; 2001US-0349575P.  
 PR 01-NOV-2001; 2001US-0346357P.  
 PR 12-APR-2002; 2002US-0371972P.  
 PR 12-APR-2002; 2002US-0371980P.  
 PR 17-APR-2002; 2002US-0373261P.  
 PR 19-APR-2002; 2002US-0373805P.  
 PR 23-APR-2002; 2002US-0374738P.  
 PR 16-MAY-2002; 2002US-0381101P.  
 PR 17-MAY-2002; 2002US-0381635P.  
 PR 29-MAY-2002; 2002US-0383830P.  
 PR 01-OCT-2002; 2002US-00262839.  
 XX  
 XX (CURA-) CURAGEN CORP.  
 XX  
 XX Alsbrook JP, Anderson DW, Boldog FL, Burgess CE, Catterton E;  
 PI Edinger SR, Ellerman K, Gerlach VL, Gorman L, Guo X, Ji W;  
 PI Kekuda R, Leach MD, Li L, Miller CE, Patturajan M, Rieger DK;  
 PI Rothenberg ME, Shimkets RA, Smithson G, Spytek KA, Taupier RJ;  
 PI Vernet CAM, Voss EZ, Zerhusen BD, Zhong M;  
 XX WPI: 2003-381625/36.  
 DR N-PSDB; ACC72173.  
 XX  
 XX NOVX polypeptides and nucleic acids useful for diagnosing, preventing or  
 PT treating NOVX-associated disorders, e.g. diabetes, obesity, cancer or  
 PT dyslipidemia, and in chromosome mapping, tissue typing or  
 PT pharmacogenomics.  
 XX  
 XX Claim 1; Page 266; 487pp; English.  
 XX  
 XX The present invention relates to novel human NOV proteins and their  
 CC coding sequences (ACC72075-ACC72181 and ABR58363-ABR58469). The NOV  
 CC proteins are useful in manufacturing a medicament for treating a syndrome  
 CC associated with a human disease. The NOV proteins and coding sequences  
 CC may be used to diagnose, treat or prevent metabolic disorders such as  
 CC diabetes or obesity, infections, cachexia, cancer, neurodegenerative  
 CC disorders such as Alzheimer's disease or Parkinson's disease, immune  
 CC disorders, haematopoietic disorders and various dyslipidaemias  
 XX  
 XX Sequence 307 AA;  
 SQ  
 Query Match 30.4%; Score 252.5; DB 6; Length 307;  
 Best Local Similarity 44.8%; Pred. No. 1.7e-17; Mismatches 24; Indels 3; Gaps 2;  
 Matches 39; Conservative 21; Mismatches 24; Indels 3; Gaps 2;  
 QY 55 DCGPGVQSVRAVWCFHVDGWTSHLSNCGESNRPKERSCFRVCDHSDLFQWEVS DMHHC 114  
 Db 4 ECGPGGIQTRAVVCAHVEGWTTLHTNCKQAEPPNQQNCFKVDWHKELYDRLGPNQC 63  
 QY 115 VLPVYARGEVKPTAEVCVTAHQGLQHR 141  
 Db 64 QPV-ISKLEKP--LECIRGEEGIQVR 87  
 RESULT 11  
 ABR58454  
 ID ABR58454 standard; protein; 1490 AA.  
 XX  
 AC ABR58454;  
 XX  
 XX 07-JUL-2003 (first entry)  
 DT Human NOV47e.  
 DE  
 XX Human; NOV; antidiabetic; anorectic; antibacterial; virucide;  
 KW immunomodulator; cytostatic; nootropic; neuroprotective; dyslipidaemia;  
 KW antiparkinsonian; antilipaeamic; gene therapy; metabolic disorder;  
 KW diabetes; obesity; infection; cachexia; cancer; Parkinson's disease;  
 KW neurodegenerative disorder; Alzheimer's disease; immune disorder;  
 KW haematopoietic disorder.



```
XX Homo sapiens.
XX WO2003029423-A2.
XX 10-APR-2003.
XX 02-OCT-2002; 2002WO-US031358.
XX 02-OCT-2001; 2001US-0326483P.
XX 05-OCT-2001; 2001US-0327342P.
XX 09-OCT-2001; 2001US-0327917P.
XX 09-OCT-2001; 2001US-0328029P.
XX 09-OCT-2001; 2001US-0328044P.
XX 12-OCT-2001; 2001US-0328056P.
XX 15-OCT-2001; 2001US-032849P.
XX 17-OCT-2001; 2001US-0330142P.
XX 22-OCT-2001; 2001US-0331058P.
XX 24-OCT-2001; 2001US-033266P.
XX 29-OCT-2001; 2001US-0343629P.
XX 01-NOV-2001; 2001US-0349575P.
XX 12-APR-2002; 2002US-0346357P.
XX 12-APR-2002; 2002US-0371972P.
XX 12-APR-2002; 2002US-0371980P.
XX 17-APR-2002; 2002US-0373261P.
XX 19-APR-2002; 2002US-0373805P.
XX 23-APR-2002; 2002US-0374738P.
XX 16-MAY-2002; 2002US-0381101P.
XX 17-MAY-2002; 2002US-0381635P.
XX 29-MAY-2002; 2002US-0383630P.
XX 01-OCT-2002; 2002US-00262839.
XX (CURA-) CURAGEN CORP.
XX Alsobrook JP, Anderson DW, Boldog FL, Burgess CE, Catterton E;
XX Edinger SR, Ellerman K, Gerlach VL, Gorman L, Guo X, Ji W;
XX Kekuda R, Leach MD, Li L, Miller CE, Patturajan M, Rieger DK;
XX Rothenberg ME, Shimkets RA, Smithson G, Spytek KA, Taupier RJ;
XX Vernet CAM, Voss EZ, Zerhusen BD, Zhong M;
XX WPI; 2003-381625/36.
XX N-PSDB; ACC72165.
XX NOVX polypeptides and nucleic acids useful for diagnosing, preventing or
XX treating NOVX-associated disorders, e.g. diabetes, obesity, cancer or
XX dyslipidemia, and in chromosome mapping, tissue typing or
XX pharmacogenomics.
XX Claim 1; Page 258-259; 487pp; English.
XX The present invention relates to novel human NOV proteins and their
XX coding sequences (ACC72075-ACC72181 and ABR58363-ABR58469). The NOV
XX proteins are useful in manufacturing a medicament for treating a syndrome
XX associated with a human disease. The NOV proteins and coding sequences
XX may be used to diagnose, treat or prevent metabolic disorders such as
XX diabetes or obesity, infections, cachexia, cancer, neurodegenerative
XX disorders such as Alzheimer's disease or Parkinson's disease, immune
XX disorders, haematopoietic disorders and various dyslipidaemias
XX Sequence 1490 AA;
XX Query Match 30.4%; Score 252.5; DB 6; Length 1490;
XX Best Local Similarity 44.8%; Freq. No. 9.7e-17;
XX Matches 39; Conservative 21; Mismatches 24; Indels 3; Gaps 2;
2Y 55 DCGPGGVQSRVAVWCFFVDGWTSHLSNCGESNRPKRCFRVCDWHSDFLQFVEVSDWHHC 114
2b 4 ECGPGGIQIRAVVCAVVEGTTLTHTNCKAERNQNCFCVCDWHKELYDRLGFWQC 63
2Y 115 VLVYPYARGEVVKPTABCTVQAQGLQHR 141
2b 64 QPV-ISKSLKFP--LECIKGEGEIQVR 87
```

```
RESULT 12
ABR58453
ID ABR58453 standard; protein; 1545 AA.
XX
AC ABR58453;
XX
DT 07-JUL-2003 (first entry)
XX
KW Human NOV47d.
XX
KW Human; NOV; antidiabetic; anorectic; antibacterial; virucide;
KW immunomodulator; cytostatic; nootropic; neuroprotective; dyslipidaemia;
KW antiparkinsonian; antilipemic; gene therapy; metabolic disorder;
KW diabetes; obesity; infection; cachexia; cancer; Parkinson's disease;
KW neurodegenerative disorder; Alzheimer's disease; immune disorder;
KW haematopoietic disorder.
XX
OS Homo sapiens.
XX
FN WO2003029423-A2.
XX
PD 10-APR-2003.
XX
PF 02-OCT-2002; 2002WO-US031358.
XX
PR 02-OCT-2001; 2001US-0326483P.
PR 05-OCT-2001; 2001US-0327342P.
PR 09-OCT-2001; 2001US-0327917P.
PR 09-OCT-2001; 2001US-0328029P.
PR 09-OCT-2001; 2001US-0328044P.
PR 09-OCT-2001; 2001US-0328056P.
PR 12-OCT-2001; 2001US-032849P.
PR 15-OCT-2001; 2001US-0329414P.
PR 17-OCT-2001; 2001US-0330142P.
PR 22-OCT-2001; 2001US-0341058P.
PR 24-OCT-2001; 2001US-033266P.
PR 29-OCT-2001; 2001US-0343629P.
PR 01-NOV-2001; 2001US-0349575P.
PR 12-APR-2002; 2002US-0346357P.
PR 12-APR-2002; 2002US-0371972P.
PR 12-APR-2002; 2002US-0371980P.
PR 17-APR-2002; 2002US-0373261P.
PR 19-APR-2002; 2002US-0373805P.
PR 23-APR-2002; 2002US-0374738P.
PR 16-MAY-2002; 2002US-0381101P.
PR 17-MAY-2002; 2002US-0381635P.
PR 29-MAY-2002; 2002US-0383630P.
PR 01-OCT-2002; 2002US-00262839.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Alsobrook JP, Anderson DW, Boldog FL, Burgess CE, Catterton E;
XX Edinger SR, Ellerman K, Gerlach VL, Gorman L, Guo X, Ji W;
XX Kekuda R, Leach MD, Li L, Miller CE, Patturajan M, Rieger DK;
XX Rothenberg ME, Shimkets RA, Smithson G, Spytek KA, Taupier RJ;
XX Vernet CAM, Voss EZ, Zerhusen BD, Zhong M;
XX WPI; 2003-381625/36.
XX N-PSDB; ACC72165.
XX NOVX polypeptides and nucleic acids useful for diagnosing, preventing or
XX treating NOVX-associated disorders, e.g. diabetes, obesity, cancer or
XX dyslipidemia, and in chromosome mapping, tissue typing or
XX pharmacogenomics.
XX Claim 1; Page 257; 487pp; English.
XX The present invention relates to novel human NOV proteins and their
XX coding sequences (ACC72075-ACC72181 and ABR58363-ABR58469). The NOV
XX proteins are useful in manufacturing a medicament for treating a syndrome
XX associated with a human disease. The NOV proteins and coding sequences
```

The present sequence is that of novel secreted proteinSCP1, which is predicted to localise in the plasma membrane. The protein shows homology to human KIAA0960 protein. RNA species with homology to the SCP1 cDNA were detected in endorheial calis, pancreas, adipose, adrenal gland, thyroid, mammary gland, myometrium, uterus, placenta, prostate testis, and in neoplastic cells derived from ovarian carcinoma, breast carcinoma, prostate carcinoma (bone metastases) and melanoma. SCP1 was also highly expressed in a small cell lung cancer, a large cell lung cancer, and a

02-OCT-2001;	2001US-0326483P.
05-OCT-2001;	2001US-0327342P.
09-OCT-2001;	2001US-0327917P.
09-OCT-2001;	2001US-0328039P.
09-OCT-2001;	2001US-0328044P.
09-OCT-2001;	2001US-0328056P.
12-OCT-2001;	2001US-0328849P.
15-OCT-2001;	2001US-0329414P.
17-OCT-2001;	2001US-0330142P.
22-OCT-2001;	2001US-0341058P.
24-OCT-2001;	2001US-0339266P.
24-OCT-2001;	2001US-0343639P.
29-OCT-2001;	2001US-0349575P.
01-NOV-2001;	2001US-0346357P.
12-APR-2002;	2002US-0371972P.
12-APR-2002;	2002US-0371980P.
17-APR-2002;	2002US-0373261P.

Search completed: February 24, 2004, 01:02:09  
Job time : 13.9894 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2004, 17:26:11 ; Search time 140.513 Seconds

(without alignments)

10566.289 Million cell updates/sec

Title: US-10-022-710-5\_COPY\_1\_424

Perfect score: 424

Sequence: 1 atgttccaaagcaacct.....gcattgacgacgacccgga 424

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 2308684 seqs, 1750822206 residues

Total number of hits satisfying chosen parameters: 4617368

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
17: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
18: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	285	67.2	744	9	US-09-864-761-27172
2	264.4	62.4	477	9	US-09-864-761-10551
3	78.6	18.5	3053	15	US-10-104-047-958
4	76	17.9	6378	14	US-10-037-270-332
5	76	17.9	6378	15	US-10-117-722-332
6	75.4	17.8	1955	9	US-09-864-761-13883
7	75.4	17.8	6373	13	US-10-000-512-1
8	75.4	17.8	6373	15	US-10-074-566-1
9	75.4	17.8	6373	15	US-10-074-566-58
10	73.8	17.4	833	9	US-09-864-761-30452
11	73.8	17.4	4810	15	US-10-074-566-40
12	35.8	8.4	1629	9	US-09-732-224-6
13	34.2	8.1	652	15	US-10-027-632-232714
14	34.2	8.1	82121	15	US-10-085-111-136
15	34.2	8.1	358246	15	US-10-292-798-1095

C 16	33.6	7.9	2939	15	US-10-080-334-69	Sequence 69, Appl
C 17	32.4	7.6	794	15	US-10-027-632-156348	Sequence 156348, App
18	32.4	7.6	1102	12	US-10-388-703-28	Sequence 28, Appl
19	32.4	7.6	1103	12	US-10-388-703-26	Sequence 26, Appl
20	32.4	7.6	1103	12	US-10-388-703-27	Sequence 27, Appl
21	32.4	7.6	1103	12	US-10-388-703-30	Sequence 30, Appl
22	32.4	7.6	2598	9	US-09-764-887-432	Sequence 432, App
C 23	32.4	7.6	2598	14	US-10-073-961-432	Sequence 432, App
C 24	32	7.5	440	14	US-10-184-644-386	Sequence 386, App
25	32	7.5	440	14	US-10-184-644-386	Sequence 386, App
26	32	7.5	440	14	US-10-063-685-134	Sequence 134, App
C 27	32	7.5	1957	14	US-10-198-846-9708	Sequence 9708, App
C 28	31.8	7.5	588	15	US-10-027-632-105925	Sequence 105925, App
C 29	31.8	7.5	588	15	US-10-027-632-131223	Sequence 131223, App
C 30	31.8	7.5	588	15	US-10-027-632-131223	Sequence 131223, App
C 31	31.8	7.5	588	15	US-10-027-632-131224	Sequence 131224, App
C 32	31.8	7.5	722	15	US-10-027-632-19326	Sequence 19326, App
C 33	31.8	7.5	32127	15	US-10-242-355-1019	Sequence 1019, App
C 34	31.4	7.4	2674	9	US-09-917-800A-1500	Sequence 1500, App
C 35	31.4	7.4	2674	15	US-10-388-934-584	Sequence 584, App
C 36	31.4	7.4	8113	14	US-10-072-900-3	Sequence 3, Appli
C 37	31.4	7.4	8350	14	US-10-072-900-3	Sequence 4, Appli
C 38	31.4	7.4	8875	14	US-10-072-900-2	Sequence 2, Appli
C 39	31.4	7.4	9112	14	US-10-072-900-1	Sequence 1, Appli
C 40	31.2	7.4	521	15	US-10-027-632-264412	Sequence 264412, App
C 41	31.2	7.4	1128	14	US-10-156-761-6651	Sequence 6651, App
C 42	31.2	7.4	203654	9	US-09-820-905-3	Sequence 3, Appli
C 43	31.2	7.4	418550	15	US-10-292-798-1463	Sequence 1463, App
C 44	31.2	7.4	9025608	14	US-10-156-761-1	Sequence 1, Appli
C 45	31	7.3	2866	10	US-09-866-050A-598	Sequence 598, App

#### ALIGNMENTS

#### RESULT 1

US-09-864-761-27172  
; Sequence 27172, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aecmca-X-1  
; CURRENT APPLICATION NUMBER: US/09/864.761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30

```
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 27172
; LENGTH: 744
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC011231.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.89
; OTHER INFORMATION: EST HUMAN HIT: BE295719.1, EVALUATE 6.00e-53
; OTHER INFORMATION: NT HIT: AF133643.1, EVALUATE 1.90e-00
; OTHER INFORMATION: SWISSPROT HIT: P35446, EVALUATE 6.00e-10
; US-09-864-761-27172

Query Match          67.2%; Score 285; DB 9; Length 744;
Best Local Similarity 100.0%; Pred. No. 2.4e-88;
Matches 285; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 140 GTCGTCGGGAAGGTGTACAGAGACTGTGGTCCCGAGGAGTCCAGAGTCGGGCAAGTGT 199
DB 1 GTCGTCGGGAAGGTGTACAGAGACTGTGGTCCCGAGGAGTCCAGAGTCGGGCAAGTGT 60
QY 200 GGTGTTTTCATGTCACGGGTGCACAGTCACCTGTCTAATCTGTGTGTGAGACACAGGC 259
DB 61 GGTGTTTTCATGTTGACGGGTGCACAGTCACCTGTCTAATCTGTGTGTGAGACACAGGC 120
QY 260 CTCCTAAGGAAGAAGTGTGTTCCGAGTTGTGACTGGCAGAGTACCTCTTTCAAGTGGG 319
DB 121 CTCCTAAGGAAGAAGTGTGTTCCGAGTTGTGACTGGCAGAGTACCTCTTTCAAGTGGG 180
QY 320 AGTTTCTGACTGGACACACTGTGTGCTTCTTACCTGCTGGGTGAAGTCAAGCCTC 379
DB 181 AGTTTCTGACTGGACACACTGTGTGCTTCTTACCTGCTGGGTGAAGTCAAGCCTC 240
QY 380 GGACTGCAGAGTGTGTGACGGCTCAGCATGGACTGCAGCACCGGA 424
DB 241 GGACTGCAGAGTGTGTGACGGCTCAGCATGGACTGCAGCACCGGA 285

RESULT 2
US-09-864-761-10551
; Sequence 10551, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
```

```
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 10551
; LENGTH: 477
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC011231.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.89
; US-09-864-761-10551

Query Match          62.4%; Score 264.4; DB 9; Length 477;
Best Local Similarity 85.3%; Pred. No. 2.9e-81;
Matches 295; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 53 GGAAGCTCTTTCTATTGCTTTCTCTTCTGCTGCCATCGAGCTCATTTGGAGGCAAAA 112
DB 132 GGAAGTGTGTAAATTGACTTCTACTTACCTGTGAATCTCTGAATCTCTGAGTAA 191
QY 113 AGGATAATCAGTTTCATCTGGAACACAGGTCGCTGGGAAAGGTGTACAGGAGACTGTGGTC 172
DB 192 TTAACATCTTGTGTTTCTGCTGTAGTCCGTGGGAGGTGTACAGGAGACTGTGGTC 251
QY 173 CCGGAGAGTCCAGAGTCGCGGAGTGTGTTTTCATGTTGACGGGTGACAAAGTCACC 232
DB 252 CCGGAGAGTCCAGAGTCGCGGAGTGTGTTTTCATGTTGACGGGTGACAAAGTCACC 311
QY 233 TGTCTAACTGTTGAGAGCAACAGAGCTCCAAAGGAAGAAGTGTGTTTCCGAGTTTGTG 292
DB 312 TGTCTAACTGTTGAGAGCAACAGAGCTCCAAAGGAAGAAGTGTGTTTCCGAGTTTGTG 371
QY 293 ACTGGCAGTGCACCTCTTTTCAGTGGAGGTTTCTGACTGGCACCACCTGTGCTCTGTTTC 352
DB 372 ACTGGCAGTGCACCTCTTTTCAGTGGAGGTTTCTGACTGGCACCACCTGTGCTCTGTTTC 431
QY 353 CTTACGCTCCGGTGAAGTCAAGCCTCGGACTCGAGAGTGTGTGAC 398
DB 432 CTTACGCTCCGGTGAAGTCAAGCCTCGGACTCGAGAGTGTGTGAC 477

RESULT 3
US-10-104-047-958
; Sequence 958, Application US/10104047
```

Publication No. US20030236392A1  
GENERAL INFORMATION:  
APPLICANT: HELIX RESEARCH INSTITUTE  
TITLE OF INVENTION: No. US20030236392A1el full length cdna  
FILE REFERENCE: HI-A0105  
CURRENT APPLICATION NUMBER: US/10/104,047  
CURRENT FILING DATE: 2002-03-25  
PRIOR APPLICATION NUMBER:  
PRIOR FILING DATE:  
NUMBER OF SEQ ID NOS: 4096  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 958  
LENGTH: 3053  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-104-047-958

Query Match 18.5%; Score 78.6; DB 15; Length 3053;  
Best Local Similarity 58.4%; Pred. No. 2.7e-16;  
Matches 157; Conservative 0; Mismatches 109; Indels 3; Gaps 1;  
2Y 90 TGCAGCTCATTTGGAGGCAAAAGGATATCATCTGGAACCAAGTCCGTGGGG 149  
Db 393 TGGGGCGCAGGGGAGGCGGCCACCTCTATCTGTGAAGACTGGTCCATGGGG 452  
2Y 150 AAGGTGTACAGGA--GACTGTGTCCCGAGGAGTCCAGAGTCCGAGCTGGGTGTTT 206  
Db 453 CCGATGTATGGAGATGAATGTGGTCCCGAGGATCCAAACGAGGGCTGTGGGTGTC 512  
2Y 207 TCATGTTCAGGGTGGCAAGTCACTGTCTAACTGTGTGAGAGCAACAGGGCTCCAAA 266  
Db 513 TCATGTGGAGGATGACTACACTGCATCTAACTGTAAAGCAGCGCCGAGAGCCCAATA 572  
2Y 267 GGAAGAGATGTTTCCGAGTTTGTGACTGGCACAGTACCTTTTCAGTGGAGGTTTC 326  
Db 573 CCAGCAGATGTTTCAAAGTTTGCAGTTGGCAAGAGATTGTACGACTGGAGACTGGG 632  
2Y 327 TGACTGGCACCACTGTGTGTCTTCTTCTT 355  
Db 633 ACCTTGAATCGGTGTCAGGCCCGGTGATT 661

RESULT 4  
US-10-037-270-332  
Sequence 332, Application US/10037270  
Publication No. US20030104529A1  
GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
APPLICANT: Liu, Chenghua  
APPLICANT: Asundi, Vinod  
APPLICANT: Zhang, Jie  
APPLICANT: Ren, Feiyan  
APPLICANT: Chen, Rui-hong  
APPLICANT: Zhao, Qing A.  
APPLICANT: Wehrman, Tom  
APPLICANT: Xue, Aigong J.  
APPLICANT: Yang, Yonghong  
APPLICANT: Wang, Jian-Rui  
APPLICANT: Zhou, Ping  
APPLICANT: Ma, Yunqing  
APPLICANT: Wang, Duanrui  
APPLICANT: Wang, Zhiwei  
APPLICANT: Tillinghast, John  
APPLICANT: Drmanac, Radoje T.  
TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and  
FILE REFERENCE: 784CIP2B  
CURRENT APPLICATION NUMBER: US/10/037,270  
CURRENT FILING DATE: 2002-01-04  
PRIOR APPLICATION NUMBER: 09/552,317  
PRIOR FILING DATE: 2000-04-25  
PRIOR APPLICATION NUMBER: 09/488,725  
PRIOR FILING DATE: 2000-01-21

NUMBER OF SEQ ID NOS: 1104  
SOFTWARE: pt\_FL\_genes Version 1.0  
SEQ ID NO 332  
LENGTH: 6378  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (271)..(5037)  
NAME/KEY: misc feature  
LOCATION: (1)..(6378)  
OTHER INFORMATION: n = a,t,c or g  
US-10-037-270-332

Query Match 17.9%; Score 76; DB 14; Length 6378;  
Best Local Similarity 59.8%; Pred. No. 3.1e-15;  
Matches 146; Conservative 0; Mismatches 95; Indels 3; Gaps 1;  
QY 115 GATAATCAGTTCATCTCGAAACACGAGTCCGTGGGGAGGTGTACAGGA---GACTGTGGT 171  
Db 229 GGTAAATCGGAAAGCTTCTTGAGAAAGTCCATGGGGCCGATGTATGGAGATGAATGTGGT 288  
QY 172 CCGGAGGAGTCCAGAGTCGGGCGAGTGGGTGTTTTCATGTTGACGGGTGGACAAAGTCAC 231  
Db 289 CCGGAGGATCCAAACGAGGGCTGTGTGGTGTCTCATGTGGAGGATGCACTACACTG 348  
QY 232 CTGTCTAACTGTGTGAGAGCAACAGGCCTCCAAAGAAAGAGTTGTTCCGAGTTGT 291  
Db 349 CATACTAATCTGAAGCAGCGCCGAGAGACCCCAATAACACGAGAGATTGTTTCAAAGTTTC 408  
QY 292 GACTGGCACAGTCACTCTTTTCAGTGGGAGGTTTCTGACTGGCACCACTGTGTGCTTGT 351  
Db 409 GATTGGCACAAGAGTTGTACGACTGGAGACTGGGACCTTGGAAATCATGTGTCAGCCCGTG 468  
QY 352 CCTT 355  
Db 469 ATTT 472

RESULT 5  
US-10-117-722-332  
Sequence 332, Application US/10117722  
Publication No. US20030219744A1  
GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
APPLICANT: Liu, Chenghua  
APPLICANT: Asundi, Vinod  
APPLICANT: Zhang, Jie  
APPLICANT: Drmanac, Radoje T.  
TITLE OF INVENTION: No. US20030219744A1el Nucleic Acids and  
FILE REFERENCE: 784CIP2B  
CURRENT APPLICATION NUMBER: US/10/117,722  
CURRENT FILING DATE: 2002-04-04  
PRIOR APPLICATION NUMBER: 09/620,312  
PRIOR FILING DATE: 2000-07-19  
PRIOR APPLICATION NUMBER: 09/552,317  
PRIOR FILING DATE: 2000-04-25  
PRIOR APPLICATION NUMBER: 09/488,725  
PRIOR FILING DATE: 2000-01-21  
NUMBER OF SEQ ID NOS: 1104  
SOFTWARE: pt\_FL\_genes Version 1.0  
SEQ ID NO 332  
LENGTH: 6378  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (271)..(5037)  
NAME/KEY: misc feature  
LOCATION: (1)..(6378)  
OTHER INFORMATION: n = a,t,c or g



US-10-117-722-332

Query Match 17.9%; Score 76; DB 15; Length 6378;  
Best Local Similarity 59.8%; Pred. No. 3.1e-15;  
Matches 146; Conservative 0; Mismatches 95; Indels 3; Gaps 1;

QY 115 GATAATCAGTTCATCTGAAACCCAGGTCCTGGGGAAGGTGTACAGGA---GACTGTGGT 171  
DB 229 GGTAAATTGGGAAGCTTCTGAGAAAGTCCATGGGCGCATGTATCGGAGATGAATGTGGT 288  
QY 172 CCCGAGGAGTCCAGAGTCGGGCACTGTGGTGTTCATGTGTACGGGTGGAACAAGTCA 231  
DB 289 CCCGAGGAGTCCAAACAGGCGCTGTGGTGTCTCATGTGAGGAGTGAAGTCACTG 348  
QY 232 CTGTCTAACTGTGGTGTGAGAGCAACAGGCTCCAAAGGAAGAGTGTTCCTCGAGTTGT 291  
DB 349 CATACTTAAGTGAACGAGCCGAGAGACCCATTAACGAGAGATTTTCAAGATTGGC 408  
QY 292 GACTGGCAGTGCCTCTTTTCAGTGGAGGTTTCTGATGGCACTGTGTGCTTGT 351  
DB 409 GATTGGCAAAAGAGTGTACGACTGGAGACTGGGACCTTGAATCAGTGTGACCCGCTG 468  
QY 352 CCTT 355  
DB 469 ATTT 472

## RESULT 6

US-09-864-761-13883  
Sequence 13883, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
FILE REFERENCE: Aemica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21

PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 13883  
LENGTH: 1956  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC004614.1  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.9  
US-09-864-761-13883

Query Match 17.8%; Score 75.4; DB 9; Length 1956;  
Best Local Similarity 62.0%; Pred. No. 2.9e-15;  
Matches 137; Conservative 0; Mismatches 81; Indels 3; Gaps 1;  
QY 138 AGTCCGTGGGAAGGTGTACAGGA---GACTGTGTGTCCTGGAGGAGTCCAGAGTCGGGC 194  
DB 189 AGTCCATGGGCGCATGTATGGAGATGAATGTGTCTCCGGAGGCATCCAAACGAGGC 248  
QY 195 AGTGTGTGTTTTCATGTGACGGGTGGACAAGTCACTGTCTTAAGTGTGGTGAGACNA 254  
DB 249 TGTGTGTGTCTCATGTGGAGGATGGACTACCTGTCATTAAGTGTAAAGCGCCGA 308  
QY 255 CAGGCTCCAAAGGAAGAGTGTTCCTGAGTTGTGCTGGCAGAGTCACTCTTTCA 314  
DB 309 GAGACCAATNACCAGCAGAAATTTTCAAGTTTTCGATTGGCACAAGAGTTGTACGA 368  
QY 315 GTGGAGGTTTCTGACTGGCACACACTGTGTGTGTTCCTT 355  
DB 369 CTGGAGACTGGGACCTTGAATCAAGTGTGAGCCCGTGATT 409

## RESULT 7

US-10-000-512-1  
Sequence 1, Application US/10000512  
Publication No. US2002016499A1  
GENERAL INFORMATION:  
APPLICANT: Shimketa, Richard A  
APPLICANT: Fernandes, Elma  
TITLE OF INVENTION: POLYPEPTIDES AND POLYNUCLEOTIDES ENCODING SAME  
FILE REFERENCE: 15966-556  
CURRENT APPLICATION NUMBER: US/10/000,512  
CURRENT FILING DATE: 2001-10-23  
PRIOR APPLICATION NUMBER: 09/619,252  
PRIOR FILING DATE: 2000-07-19  
PRIOR APPLICATION NUMBER: 60/167,785  
PRIOR FILING DATE: 1999-11-29  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 6373  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (235)..(4998)  
NAME/KEY: misc feature  
LOCATION: (6349)  
OTHER INFORMATION: Wherein n is a or t or g or c.  
US-10-000-512-1

Query Match 17.8%; Score 75.4; DB 13; Length 6373;  
Best Local Similarity 62.0%; Pred. No. 4.9e-15;  
Matches 137; Conservative 0; Mismatches 81; Indels 3; Gaps 1;  
QY 138 AGTCCGTGGGAAGGTGTACAGGA---GACTGTGTGTCCTGGAGGAGTCCAGAGTCGGGC 194  
DB 216 AGTCCATGGGCGCATGTATGGAGATGAATGTGTCTCCGGAGGCATCCAAACGAGGC 275

QY 195 AGTGTGTTGTTTTCATGTTGACGGTGGCAAGTCACTGTCTAACTGTGGTGAGAGCAA 254  
DB 276 TGTGTGTGTGCTCATGTGGAGGATGATCACTGCTACTACTGTTAGCAGGCCGA 335  
QY 255 CAGGCTCCAAAGAAAGATGTTTCCGAGTTTGTGACTGGCAGTCACTCTTTCA 314  
DB 336 GAGACCAATAACACGAGAGATTTGTTTCAAAGTTTGGATTGGCAAAAGAGTTGTACGA 395  
QY 315 GTGGAGGTTTCTGACTGGCAGCAGTGTGCTGTGTTTCCTT 355  
DB 396 CTGGAGACTGGACCTTGGATCATGTTGACCCGCTGATT 436

RESULT 8  
US-10-074-566-1  
; Sequence 1, Application US/10074566  
; Publication No. US20030207348A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimkets, Richard A.  
; APPLICANT: Fernandes, Elma R.  
; APPLICANT: Li, Li  
; APPLICANT: Gorman, Linda  
; APPLICANT: Gusev, Vladimir Y.  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Shenoy, Suresh G.  
; APPLICANT: Spytek, Kimberly A.  
; TITLE OF INVENTION: Polypeptides and Polynucleotides Encoding Same  
; FILE REFERENCE: 15966-556 CIP1  
; CURRENT APPLICATION NUMBER: US/10/074,566  
; CURRENT FILING DATE: 2002-02-13  
; PRIOR APPLICATION NUMBER: 09/619,252  
; PRIOR FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 60/144,722  
; PRIOR FILING DATE: 1999-07-20  
; PRIOR APPLICATION NUMBER: 60/167,785  
; PRIOR FILING DATE: 1999-11-29  
; PRIOR APPLICATION NUMBER: 60/276,994  
; PRIOR FILING DATE: 2001-03-19  
; PRIOR APPLICATION NUMBER: 60/280,898  
; PRIOR FILING DATE: 2001-04-02  
; PRIOR APPLICATION NUMBER: 60/332,241  
; PRIOR FILING DATE: 2001-11-14  
; PRIOR APPLICATION NUMBER: 60/288,062  
; PRIOR FILING DATE: 2001-05-02  
; PRIOR APPLICATION NUMBER: 60/291,766  
; PRIOR FILING DATE: 2001-08-21  
; NUMBER OF SEQ ID NOS: 132  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 6373  
; TYPE: DNA  
; ORGANISM: human  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (6349)  
; OTHER INFORMATION: Wherein N is A, or T, or C, or G.

Query Match 17.8%; Score 75.4; DB 15; Length 6373;  
Best Local Similarity 62.0%; Pred. No. 4.9e-15;  
Matches 137; Conservative 0; Mismatches 81; Indels 3; Gaps 1;  
QY 138 AGTGTGTTGTTTTCATGTTGACGGTGGCAAGTCACTGTCTAACTGTGGTGAGAGCAA 194  
DB 216 AGTGTGTTGTTTTCATGTTGACGGTGGCAAGTCACTGTCTAACTGTGGTGAGAGCAA 254  
QY 195 AGTGTGTTGTTTTCATGTTGACGGTGGCAAGTCACTGTCTAACTGTGGTGAGAGCAA 254  
DB 276 TGTGTGTGTGCTCATGTGGAGGATGATGTTTCCGAGTTTGTGACTGGCAGTCACTCTTTCA 314  
QY 315 GTGGAGGTTTCTGACTGGCAGCAGTGTGCTGTGTTTCCTT 355  
DB 396 CTGGAGACTGGACCTTGGATCATGTTGACCCGCTGATT 436

QY 255 CAGGCTCCAAAGAAAGATGTTTCCGAGTTTGTGACTGGCAGTCACTCTTTCA 314  
DB 336 GAGACCAATAACACGAGAGATTTGTTTCAAAGTTTGGATTGGCAAAAGAGTTGTACGA 395  
QY 315 GTGGAGGTTTCTGACTGGCAGCAGTGTGCTGTGTTTCCTT 355  
DB 396 CTGGAGACTGGACCTTGGATCATGTTGACCCGCTGATT 436

RESULT 9  
US-10-074-566-58  
; Sequence 58, Application US/10074566  
; Publication No. US20030207348A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimkets, Richard A.  
; APPLICANT: Fernandes, Elma R.  
; APPLICANT: Li, Li  
; APPLICANT: Gorman, Linda  
; APPLICANT: Gusev, Vladimir Y.  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Shenoy, Suresh G.  
; APPLICANT: Spytek, Kimberly A.  
; TITLE OF INVENTION: Polypeptides and Polynucleotides Encoding Same  
; FILE REFERENCE: 15966-556 CIP1  
; CURRENT APPLICATION NUMBER: US/10/074,566  
; CURRENT FILING DATE: 2002-02-13  
; PRIOR APPLICATION NUMBER: 09/619,252  
; PRIOR FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 60/144,722  
; PRIOR FILING DATE: 1999-07-20  
; PRIOR APPLICATION NUMBER: 60/167,785  
; PRIOR FILING DATE: 1999-11-29  
; PRIOR APPLICATION NUMBER: 60/276,994  
; PRIOR FILING DATE: 2001-03-19  
; PRIOR APPLICATION NUMBER: 60/280,898  
; PRIOR FILING DATE: 2001-04-02  
; PRIOR APPLICATION NUMBER: 60/332,241  
; PRIOR FILING DATE: 2001-11-14  
; PRIOR APPLICATION NUMBER: 60/288,062  
; PRIOR FILING DATE: 2001-05-02  
; PRIOR APPLICATION NUMBER: 60/291,766  
; PRIOR FILING DATE: 2001-08-21  
; NUMBER OF SEQ ID NOS: 132  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 58  
; LENGTH: 6373  
; TYPE: DNA  
; ORGANISM: human  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (6349)  
; OTHER INFORMATION: Wherein N is A, or T, or C, or G.

Query Match 17.8%; Score 75.4; DB 15; Length 6373;  
Best Local Similarity 62.0%; Pred. No. 4.9e-15;  
Matches 137; Conservative 0; Mismatches 81; Indels 3; Gaps 1;  
QY 138 AGTGTGTTGTTTTCATGTTGACGGTGGCAAGTCACTGTCTAACTGTGGTGAGAGCAA 194  
DB 216 AGTGTGTTGTTTTCATGTTGACGGTGGCAAGTCACTGTCTAACTGTGGTGAGAGCAA 254  
QY 195 AGTGTGTTGTTTTCATGTTGACGGTGGCAAGTCACTGTCTAACTGTGGTGAGAGCAA 254  
DB 276 TGTGTGTGTGCTCATGTGGAGGATGATGTTTCCGAGTTTGTGACTGGCAGTCACTCTTTCA 314  
QY 315 GTGGAGGTTTCTGACTGGCAGCAGTGTGCTGTGTTTCCTT 355  
DB 396 CTGGAGACTGGACCTTGGATCATGTTGACCCGCTGATT 436

QY 315 GTGGAGCTTTCTGACTGGCACCACCTGTGCTGTTCCTT 355  
DB 336 CTGGAGACTGGGACCTTGGAAATCAGTGTGACCCCGTGATT 436

RESULT 10  
US-09-864-761-30452  
; Sequence 30452, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aemica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 4917  
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 30452  
; LENGTH: 833  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC004614.1  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.9  
; OTHER INFORMATION: SWISSPROT HIT: P35446, EVALUE 3.00e-12  
; OTHER INFORMATION: NT HIT: ABC30073.1, EVALUE 1.10e-01  
; OTHER INFORMATION: EST\_HUMAN HIT: AW612526.1, EVALUE 0.00e+00  
US-09-864-761-30452  
Query Match 17.4%; Score 73.8; DB 9; Length 833;  
Best Local Similarity 60.9%; Pred. No. 7.2e-15;  
Matches 120; Conservative 0; Mismatches 77; Indels 0; Gaps 0;  
QY 159 AGGAGACTGTGTCCCGGAGGAGTCCAGAGTCGGCAGTGTGTTCATTTTCATTTGACGG 218  
DB 23 AGATGAATGTGTCCCGGAGGCATCCAAACGAGGCGCTGTGTGTCTCATGTGGAGGG 82

Matches 120; Conservative 0; Mismatches 77; Indels 0; Gaps 0;  
QY 159 AGGAGACTGTGTCCCGGAGGAGTCCAGAGTCGGCAGTGTGTTCATTTTCATTTGACGG 218  
DB 23 AGATGAATGTGTCCCGGAGGCATCCAAACGAGGCGCTGTGTGTCTCATGTGGAGGG 82  
QY 219 GTGGACAAGTCACTGTCTAACTGTGTGAGAGCAACAGGCCTCCAAAGGAAGAAGTTG 278  
DB 83 ATGGACTACACTGCATACACTGTGTAAGCAGCGCCGAGAGACCAATAACCCAGACAATTG 142  
QY 279 TTTCGAGTTTGTGACTGCCACAGTGACCTCTTTAGTGGGAGGTTTCTGACTGGCACCA 338  
DB 143 TTTCAAAAGTTTCGATTGTCACAAAAGAGTTGACGACTGGGACCTTGGAAATCA 202  
QY 339 CTGTGTGCTTGTTCCTT 355  
DB 203 GTGTACGCCCGTGATT 219

RESULT 11  
US-10-074-566-40  
; Sequence 40, Application US/10074566  
; Publication No. US20030207348A1  
; GENERAL INFORMATION:  
; APPLICANT: Shinkets, Richard A.  
; APPLICANT: Fernandes, Elma R.  
; APPLICANT: Li, Li  
; APPLICANT: Gorman, Linda  
; APPLICANT: Gusev, Vladimir Y.  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Shenoy, Suresh G.  
; APPLICANT: Spytek, Kimberly A.  
; TITLE OF INVENTION: Polypeptides and Polynucleotides Encoding Same  
; FILE REFERENCE: 15966-556 CIP1  
; CURRENT APPLICATION NUMBER: US/10/074,566  
; CURRENT FILING DATE: 2002-02-13  
; PRIOR APPLICATION NUMBER: 09/619,252  
; PRIOR FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 60/144,722  
; PRIOR FILING DATE: 1999-07-20  
; PRIOR APPLICATION NUMBER: 60/167,785  
; PRIOR FILING DATE: 1999-11-29  
; PRIOR APPLICATION NUMBER: 60/276,994  
; PRIOR FILING DATE: 2001-03-19  
; PRIOR APPLICATION NUMBER: 60/280,898  
; PRIOR FILING DATE: 2001-04-02  
; PRIOR APPLICATION NUMBER: 60/332,241  
; PRIOR FILING DATE: 2001-11-14  
; PRIOR APPLICATION NUMBER: 60/288,062  
; PRIOR FILING DATE: 2001-05-02  
; PRIOR APPLICATION NUMBER: 60/291,766  
; PRIOR FILING DATE: 2001-05-17  
; PRIOR APPLICATION NUMBER: 60/314,007  
; PRIOR FILING DATE: 2001-08-21  
; NUMBER OF SEQ ID NOS: 132  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 40  
; LENGTH: 4810  
; TYPE: DNA  
; ORGANISM: human  
; US-10-074-566-40

Query Match 17.4%; Score 73.8; DB 15; Length 4810;  
Best Local Similarity 60.9%; Pred. No. 1.6e-14;  
Matches 120; Conservative 0; Mismatches 77; Indels 0; Gaps 0;  
QY 159 AGGAGACTGTGTCCCGGAGGAGTCCAGAGTCGGCAGTGTGTTCATTTTCATTTGACGG 218  
DB 23 AGATGAATGTGTGTCCCGGAGGCATCCAAACGAGGCGCTGTGTGTCTCATGTGGAGGG 82  
QY 219 GTGGACAAGTCACTGTCTAACTGTGTGAGAGCAACAGGCCTCCAAAGGAAGAAGTTG 278

```

; 83 ATGACTACACTGCATCTAATCTAAGTGAAGCCGCGAGAGACCAATAACCCAGCAATG 142
; 279 TTTCGAGTTTGACTGGACAGTCACTCTTTCACTGGGAGTTTCTGATCGGCACCA 338
; 143 TTTCGAGTTTGACTGGACAGTCACTCTTTCACTGGGAGTTTCTGATCGGCACCA 202
; 339 CTGTGCTGCTGTTCCTT 355
; 203 GTGTCAAGCCGTGATTT 219

RESULT 12
JS-09-732-224-6
; Sequence 6, Application US/09732224
; Patent No. US20020095022A1
; GENERAL INFORMATION:
; APPLICANT: Holloway, James L.
; TITLE OF INVENTION: Human Secretin-like Receptor
; FILE REFERENCE: 99-100
; CURRENT APPLICATION NUMBER: US/09/732,224
; CURRENT FILING DATE: 2000-12-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1629
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: This degenerate nucleotide sequence encodes the
; OTHER INFORMATION: amino acid sequence of SEQ ID NO:5.
; NAME/KEY: variation
; LOCATION: (1)...(1629)
; OTHER INFORMATION: N is any nucleotide.
JS-09-732-224-6

Query Match      8.4%; Score 35.8; DB 9; Length 1629;
Best Local Similarity 28.9%; Pred. No. 0.16; Indels 0; Gaps 0;
Matches 65; Conservative 35; Mismatches 125;

; 23 CAGTCACTTGTGGTATGAGGAGCATGAGAAAGCTCTTCTATTTCTCTCTTGC 82
; 512 CNGTAAATWNSNTTGTGCAAYACARWSNTYNTNGNGNCNCNCTTTC 571
; 83 TGTCCTCATGCTCATTTGGAAGGCAAAAGATAATCAGTTCATCTGGAACAGGTC 142
; 572 CNYTNCAYCCTGCCNWSNGARGGNTAYACNYTNACNTGYGNTNTTYTGGARGGGNG 631
; 143 CGTGGGGAAGGTGTACAGAGACTGTGGTCCCGGAGAGTCCAGAGTCCGGCAGTGTGT 202
; 632 CNGMNAARCARCCTCGGNGNGTGSNCCNGARGGNTGYMGNACNGARCCCNWSNC 691
; 203 GTTTTCATGTGACGGGTGACAAAGTCACTGTCTTAACGTGTGTG 247
; 692 AYWSNCARGNTNTYNTGYMGTGYAAYCAVYTNACNTAYTGTGNG 736

RESULT 13
JS-10-027-632-232714/c
; Sequence 232714, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS in the Human Genome
; FILE REFERENCE: 108927.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29

; 83 ATGACTACACTGCATCTAATCTAAGTGAAGCCGCGAGAGACCAATAACCCAGCAATG 142
; 279 TTTCGAGTTTGACTGGACAGTCACTCTTTCACTGGGAGTTTCTGATCGGCACCA 338
; 143 TTTCGAGTTTGACTGGACAGTCACTCTTTCACTGGGAGTTTCTGATCGGCACCA 202
; 339 CTGTGCTGCTGTTCCTT 355
; 203 GTGTCAAGCCGTGATTT 219

; 77 TCTTGCTGTCCATGTCAGCTCATTTGGAAGGCAAAAGGATAATCAGTTCATCTGGAAC 136
; 480 TATGATGCTCTAGACAGCAATGTAGCATTTCCCAAGGATGAGTGTGCCACTTCAGCT 421
; 137 CAGTCCGTGGGGAAGGTGTACAGGAGACTGTGTCGCCGAGAGTCCAGAGTCGGGCG 196
; 420 CAGCCTGGAGGGGCGGACAGTCACTGGAAGGATGAGTGTGAGTGTGAGTGTCTACAG 361
; 197 TGTGCTGTCTTCTATGTTGACGGGTGGCAAGTCACTGTCTTAACGTGTGTCAGAGCA 256
; 360 TAGCATGCGCCGATGGGTAGGTGTAGCAGCAGCTCTTCTCAGGATGTCAGACCA 301
; 257 GGC 259
; 300 GGC 298

RESULT 14
US-10-085-117-136/c
; Sequence 136, Application US/10085117
; Publication No. US200302334A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: 529452000121
; CURRENT APPLICATION NUMBER: US/10/085,117
; CURRENT FILING DATE: 2002-02-27
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 136
; LENGTH: 82121
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: variation
; LOCATION: (1)...(82121)
; OTHER INFORMATION: n = any nucleotide
US-10-085-117-136

Query Match      8.1%; Score 34.2; DB 15; Length 82121;
Best Local Similarity 49.2%; Pred. No. 3.3;
Matches 90; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

; 165 CTGTGCTCCCGAGGAGTCCAGATCGGCGGAGTGTGTGTTTCTGATCTTACCGGTGAC 224
; 74846 CCGGGGTCTCTCTGGTCTGATGATGATGATGATGATGATGATGATGATGATGATG 74787
```

QY 225 AAGTCACCTGTCTACTGTGTGAGCAACAGCGCTCCAAAGGAAGAAGTGTTCCTCCG 284  
Db 74786 TGGGACACCTGTGTGAGCTTTGTGGGAAGAGGGCTTGGGGCCCTGTGGAGGCTGCC 74727  
QY 285 AGTTTGTGACTGGCAGCTGACCTCTTTTCACTGGGAGGTTTCTGACTGGCACCAGCTGTGT 344  
Db 74726 CAGGGGTGACTGCCACGGTGTCTTCTTTCCAGGGAACCTTCTGACCAGCCTCATGGGCT 74667  
QY 345 GCT 347  
Db 74666 CCT 74664

## RESULT 15

US-10-292-798-1095/c  
; Sequence 1095, Application US/10292798  
; Publication No. US20030235833A1  
; GENERAL INFORMATION:  
; APPLICANT: SUWA, MAKIKO  
; APPLICANT: ASAI, KIYOSHI  
; APPLICANT: AKIYAMA, YUTAKA  
; APPLICANT: ABURATANI, HIROYUKI  
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS  
; FILE REFERENCE: 084335/166  
; CURRENT APPLICATION NUMBER: US/10/292,798  
; CURRENT FILING DATE: 2002-11-13  
; PRIOR APPLICATION NUMBER: 10/017,161  
; PRIOR FILING DATE: 2001-12-18  
; PRIOR APPLICATION NUMBER: JP 2001-246789  
; PRIOR FILING DATE: 2001-06-18  
; NUMBER OF SEQ ID NOS: 2070  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1095  
; LENGTH: 358246  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; LOCATION: source  
; FEATURE: (1) .. (358246)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (201) .. (326)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (605) .. (1294)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (29165) .. (29206)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (39051) .. (39221)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (72539) .. (72763)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (133049) .. (133353)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (261460) .. (261526)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (291173) .. (291325)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (291513) .. (291770)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (302538) .. (302800)  
; FEATURE:  
; NAME/KEY: CDS

; LOCATION: (325058) .. (325316)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (330441) .. (330609)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (337750) .. (337828)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (343822) .. (344071)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (358008) .. (358046)  
; FEATURE:  
; NAME/KEY: modified base  
; LOCATION: (547) .. (546)  
; OTHER INFORMATION: a, t, c, g, unknown or other  
; FEATURE:  
; NAME/KEY: modified base  
; LOCATION: (1659) .. (1758)  
; OTHER INFORMATION: a, t, c, g, unknown or other  
; FEATURE:  
; NAME/KEY: modified base  
; LOCATION: (2818) .. (2917)  
; OTHER INFORMATION: a, t, c, g, unknown or other  
; FEATURE:  
; NAME/KEY: modified base  
; LOCATION: (1719) .. (17278)  
; OTHER INFORMATION: a, t, c, g, unknown or other  
; FEATURE:  
; NAME/KEY: modified base  
; LOCATION: (17497) .. (17497)  
; OTHER INFORMATION: a, t, c, g, unknown or other  
; FEATURE:  
; NAME/KEY: modified base  
; LOCATION: (17499) .. (17499)  
; OTHER INFORMATION: a, t, c, g, unknown or other  
; FEATURE:  
; NAME/KEY: modified base  
; LOCATION: (17518) .. (17518)  
; OTHER INFORMATION: a, t, c, g, unknown or other  
; FEATURE:  
; NAME/KEY: modified base  
; LOCATION: (51382) .. (51481)  
; OTHER INFORMATION: a, t, c, g, unknown or other  
; FEATURE:  
; NAME/KEY: modified base  
; LOCATION: (64360) .. (64459)  
; OTHER INFORMATION: a, t, c, g, unknown or other  
; FEATURE:  
; NAME/KEY: modified base  
; LOCATION: (71021) .. (71120)  
; OTHER INFORMATION: a, t, c, g, unknown or other  
; FEATURE:  
; NAME/KEY: modified base  
; LOCATION: (77979) .. (78078)  
; OTHER INFORMATION: a, t, c, g, unknown or other  
; FEATURE:  
; NAME/KEY: modified base  
; LOCATION: (81210) .. (81309)  
; OTHER INFORMATION: a, t, c, g, unknown or other  
; FEATURE:  
; NAME/KEY: modified base  
; LOCATION: (103441) .. (103540)  
; OTHER INFORMATION: a, t, c, g, unknown or other  
; FEATURE:  
; NAME/KEY: modified base  
; LOCATION: (130963) .. (130964)  
; OTHER INFORMATION: a, t, c, g, unknown or other  
; FEATURE:  
; NAME/KEY: modified base  
; LOCATION: (130976) .. (130977)  
; OTHER INFORMATION: a, t, c, g, unknown or other

```

FEATURE:
NAME/KEY: modified_base
LOCATION: (130979)..(130979)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (130984)..(130984)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (130987)..(130987)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (130997)..(130999)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (131002)..(131002)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (131019)..(131020)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (131022)..(131022)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (131025)..(131025)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (131031)..(131031)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (131034)..(131034)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (131048)..(131049)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (131059)..(131060)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (131063)..(131063)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (131065)..(131065)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (131070)..(131070)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (131090)..(131091)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (131127)..(131129)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base

```

Query Match 8.1%; Score 34.2; DB 15; Length 358246;  
 Best Local Similarity 49.2%; Pred. No. 6.3;

```

Matches 90; Conservative 0; Mismatches 93; Indels 0; Gaps 0;
QY 77 TCTTCTCTCCATGCAGCTCATTTGGAGGCAAAAGGATAATCAGTTTCATCTGGAAC 136
Db 352335 TATGGATGCTCTAGACAGCCCAATGTAGCAATTTCCCGAGGATGCCAGTGTGCCACTTCAGCT 352276
QY 137 CAGGTCCGTGGGGAAGGTGTACAGGAGACTGTGTCCCGAGGAGTCCAGAGTCGGGCGAG 196
Db 352275 CAGGCTCGAGGGGCGAGGCGACAGTAGCAACTGGAGTGGTAGATGAAGTAGTTCTACAG 352216
QY 197 TGTGTGTTTTTCATGTTGACGGGTGGACAAAGTCACTGTCTTAAGTGTGGTGAGAGCAACA 256
Db 352215 TAGCATGGCCCGCATGGGGTAGGGTGTAGCAGAGCTCTTCTCAGGGATGGGAGACCACCA 352156
QY 257 GGC 259
Db 352155 GGC 352153

```

Search completed: February 24, 2004, 00:59:52  
 Job time : 149.513 secs



GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

XM protein - protein search, using sw model

run on: February 24, 2004, 00:32:36 ; Search time 4.94835 Seconds  
(without alignments)  
2760.353 Million cell updates/sec

Title: US-10-022-710-6\_COPY\_1\_142

Perfect score: 830

Sequence: 1 MPPKSNLTVTCWWSMRKL.....EVKPTAEVCVTAQHGLQHRM 142

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:\*\*

1: PIR1:\*\*

2: PIR2:\*\*

3: PIR3:\*\*

4: PIR4:\*\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	119.5	14.4	951	2 T00260	hypothetical prote
2	117.5	14.2	898	2 T14764	hypothetical prote
3	114.5	13.8	1059	2 T22545	hypothetical prote
4	114.5	13.7	951	2 T00017	gene ADAMTS-1 prot
5	109.5	13.2	2165	2 T21371	hypothetical prote
6	107.5	13.0	1205	2 T18517	procollagen N-endo
7	103	12.4	1558	2 C89114	protein C37C3.6a l
8	103	12.4	2167	2 T34395	hypothetical prote
9	92	11.1	550	2 T47158	hypothetical prote
10	90.5	10.9	111	2 T34565	hypothetical prote
11	88	10.6	1184	2 T09484	cartilage intermed
12	85	10.2	142	2 JT0573	retinoic acid-indu
13	82	9.9	1444	2 T18856	angiogenesis inhib
14	81.5	9.8	1497	2 T25061	procollagen type V
15	79.5	9.6	788	2 T19576	hypothetical prote
16	79.5	9.6	957	2 T15376	hypothetical prote
17	79.5	9.6	1584	2 T00026	brain-specific ang
18	76.5	9.2	143	2 JH0385	midkine precursor
19	76.5	9.2	837	2 T00355	hypothetical prote
20	76	9.2	380	2 G01639	transmembrane prot
21	75.5	9.1	497	2 I38966	cytochrome P450 -
22	75.5	9.1	974	2 T04910	hypothetical prote
23	74	8.9	1496	1 CGH22V	collagen alpha 2IV
24	73.5	8.9	1074	2 JC5928	semaphorin F precu
25	73.5	8.9	1344	2 S47412	gene P2 protein -
26	72.5	8.7	232	2 S32398	serine proteinase
27	72.5	8.7	4544	1 S02392	alpha-2-macroglobu
28	71.5	8.6	443	2 T01697	omega-3 fatty acid
29	71.5	8.6	1548	2 S34583	serine proteinase

#### ALIGNMENTS

##### RESULT 1

T00260

hypothetical protein KIAA0605 - human

C:Species: Homo sapiens (man)

C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 21-Jul-2000

C:Accession: T00260

R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.

DNA Res. 5, 31-39, 1998

A:Title: Prediction of the coding sequences of unidentified human genes. IX. The complet-

A:Reference number: Z14086; MUID:98290545; PMID:9628591

A:Accession: T00260

A>Status: preliminary; translated from GE/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-951 <NAG>

A:Cross-references: EMBL:AB011177; NID:G3043733; PIDN:BAA25531.1; PID:G3043734

A:Experimental source: brain

C:Genetics:

A:Note: KIAA0605

C:Superfamily: thrombospondin type 1 repeat homology

F:46-106/Domain: thrombospondin type 1 repeat homology <THR1>

Query Match 14.4%; Score 119.5; DB 2; Length 951;

Best Local Similarity 34.1%; Pred. No. 0.00078;

Matches 30; Conservative 8; Mismatches 39; Indels 11; Gaps 4;

QY 31 AAHLEGKDNQFTWKPGWGRCTGDCGPGVQSRVWCFHVDG---WTSLSNCGSNRP 87

DB 785 AIHPCGDKNCPAHWLAQDWERCNTTCG-RGVKKRLVLCMELANGKPQTRSGPGCGLAACP 843

QY 88 PKERSCF-RVCDWHSDLFQWESVDWHHC 114

DB 844 PEESTCFERPC-----FKWYTSFWSEC 865

##### RESULT 2

T14764

hypothetical protein DKFZp434H204.1 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999

C:Accession: T14764

R:Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, August 1999

A:Reference number: Z18181

A:Accession: T14764

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-898 <WAM>

A:Cross-references: EMBL:AL110226

A:Experimental source: adult testis; clone DKFZp434H204

C:Genetics:

A:Note: DKFZp434H204.1

Query Match 14.2%; Score 117.5; DB 2; Length 898;  
Best Local Similarity 29.8%; Pred. No. 0.0012;  
Matches 31; Conservative 4; Mismatches 42; Indels 27; Gaps 4;

QY 44 WKPGWGRCTGCGGGVQSVRAVWCFHVDGWTSHLS-----NCGESNRPKRSR--VC 97  
DB 95 WAGWQVLCSSCGPGLSRAVLIRSVGLDEQSALEPPACELHPRPTTETPCNRHVC 154  
QY 98 DWHSDFQWEYSVDHWHCVLVPYARGVKPRVACVTAQGLQHR 141  
DB 155 P-----ATWAVGNWSQC-----SVTCGEGTQRR 177

RESULT 3  
T22545  
hypothetical protein F53B6.2 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T22545  
R:White, S.  
submitted to the EMBL Data Library, October 1996  
A:Reference number: Z19578  
A:Accession: T22545  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1059 <WIL>  
A:Cross-references: EMBL:Z81086; PIDN:CAB03121.1; GSPDB:GN00019; CESP:F53B6.2  
A:Experimental source: clone F53B6  
C:Genetics:  
A:Gene: CESP:F53B6.2  
A:Map position: 1  
A:Introns: 38/3; 92/3; 131/3; 159/3; 236/2; 282/3; 349/2; 376/2; 420/1; 576/3; 759/3; 84

Query Match 13.8%; Score 114.5; DB 2; Length 1059;  
Best Local Similarity 28.6%; Pred. No. 0.0027;  
Matches 34; Conservative 10; Mismatches 32; Indels 43; Gaps 8;

QY 44 WKPGWGRCTGCGGGVQSVRAVWCFHVDGWTSHLSNCGESNRPKRSR--VC 94  
DB 296 WEAGKSKCTASCGO-GVRRHVAC--VGG-----SDCEGRPRQETTCVAGIPCSIAT 347

QY 95 RVCDW-----HSDLFOWE-----VSDWHCVLVPYARGVKPRVACV 132  
DB 348 NSLDWNDRAYLDGNTFGSMNDHND--WQAPRLVAGENWSTCSST--CGTGVMSRTVECV 401

RESULT 4  
T00017  
Gene ADAMTS-1 protein - mouse  
C:Species: Mus musculus (house mouse)  
C>Date: 22-Jan-1999 #sequence\_revision 22-Jan-1999 #text\_change 21-Jul-2000  
C:Accession: T00017  
R:Kuno, K.; Lizasa, H.; Ohno, S.; Matsushima, K.  
Genomics 46, 466-471, 1997  
A:Title: The exon/intron organization and chromosomal mapping of the mouse ADAMTS-1 gene  
A:Reference number: Z14055; MUID:98110583; PMID:9441751  
A:Accession: T00017  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-951 <KUN>  
A:Cross-references: EMBL:AB001735; NID:g2809056; PIDN:BA244501.1; PID:g2809057  
A:Experimental source: strain 129SVJ  
C:Genetics:  
A:Gene: ADAMTS-1  
A:Introns: 228/1; 343/3; 388/1; 444/1; 539/3; 602/1; 660/3; 719/2  
C:Superfamily: thrombospondin type 1 repeat homology  
F:542-598/Domain: thrombospondin type 1 repeat homology <THR3>

Query Match 13.7%; Score 114; DB 2; Length 951;  
Best Local Similarity 28.7%; Pred. No. 0.0027;  
Matches 37; Conservative 15; Mismatches 49; Indels 28; Gaps 8;

QY 26 LLLSHA-----AHLEGGKQNF-----TWKPGWGRCTGCGGGVQSVRAVWCFH 71

Db 810 LVNGHALRPKIKFTYFMKKKTESFNAITPSEWTEENGESKTCG-SGWQRVVQCRDI 868  
QY 72 DQWTSHLNCGESNRPKRSR--VC-----PQWESVDHWHCVLVPYARGVKPETA 129  
Db 869 NGHPA--SECAKEVPKSTRPC-----ADLPCPHWQVGDWSPCKT-CGKG-YKKRTL 917  
QY 130 ECVTAQHGL 138  
Db 918 KCVSHDGGV 926

RESULT 5  
T21371  
hypothetical protein F25H8.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T21371; T24896  
R:Gajadaty, S.  
submitted to the EMBL Data Library, February 1996  
A:Reference number: Z19413  
A:Accession: T21371  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2165 <WIL>  
A:Cross-references: EMBL:Z69360; PIDN:CAA93287.1; GSPDB:GN00022; CESP:F25H8.3  
A:Experimental source: clone F25H8  
R:Gajadaty, S.  
submitted to the EMBL Data Library, February 1996  
A:Reference number: Z19949  
A:Accession: T24896  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2165 <WIL>  
A:Cross-references: EMBL:Z69361; PIDN:CAA93288.1; GSPDB:GN00022; CESP:F25H8.3  
A:Experimental source: clone T13H10  
C:Genetics:  
A:Gene: CESP:F25H8.3  
A:Map position: 4  
A:Introns: 33/1; 52/1; 135/2; 193/3; 216/1; 266/1; 495/2; 547/3; 584/3; 634/2; 744/1; 81

Query Match 13.2%; Score 109.5; DB 2; Length 2165;  
Best Local Similarity 28.6%; Pred. No. 0.015;  
Matches 28; Conservative 13; Mismatches 32; Indels 25; Gaps 4;

QY 1 MPKSNLTVCWVWSMRKFLLLSLLSHAHLSGKXDNQFIWKPGWGRCTGCGGG 60  
Db 1865 LPFKPNESTCSL-----NPDSEFKWFGPWGECCKXCGQ-G 1901

QY 61 VOSRAVWCFHVDGWTSHLSNCGESNRPKRSR--VC 97  
Db 1902 IRRRRVKCVANDGRVVRVKC-TTKKPRTOYCFERN 1938

RESULT 6  
T18517  
procollagen N-endopeptidase (EC 3.4.24.14) I - bovine  
N:Alternate names: procollagen N-proteinase  
C:Species: Bos primigenius taurus (cattle)  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T18517  
R:Colige, A.; Nusgens, B.V.; Lapiere, C.M.  
submitted to the EMBL Data Library, February 1996  
A:Description: Cloning of the cDNA of the bovine procollagen I N-proteinase.  
A:Reference number: Z18941  
A:Accession: T18517  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1205 <COL>  
A:Cross-references: EMBL:X96389; NID:e990769; PID:e228215; PIDN:CAA65253.1  
A:Experimental source: skin  
C:Genetics:  
A:Gene: PC I-NP

C:Function:  
A:Description: catalyzes cleavage of the propeptides of type I and II collagens prior to  
C:Keywords: hydrolase; metalloproteinase

Query Match 13.0%; Score 107.5; DB 2; Length 1205;  
Best Local Similarity 28.0%; Pred. No. 0.014;  
Matches 30; Conservative 13; Mismatches 37; Indels 27; Gaps 6;  
2Y 40 NQFIKPGWGRCTGCGPGVQSRVAVWC---PHVDGWTSLHSLNCGESNRPPKERSCFR 95  
DB 908 SQPVVWTGWEPCSGRGTMQVRVRCVQPLHNTTTSVHTKHCNDA-REGGRACNR 966  
2Y 96 -VCDWHSDLFOWEVSQVHCHVLPYARGEVKPTAEVTAQGLQHR 141  
DB 967 ELCPG-----RWRAGWSQC-----SVTCGNGTQER 992

## RESULT 7

89114  
protein C37C3.6a [imported] - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 10-May-2001  
C:Accession: C89114  
R:anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog  
A:Reference number: A75000; PMID:99069613; PMID:9851916  
A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C.ele  
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
A:Accession: C89114  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1558 <STO>  
A:Cross-references: GB:chr\_V; PIDN:AAC25867.1; PID:G3294501; GSPDB:GN00023; CESP:C37C3.6  
C:Genetics:  
A:Gene: C37C3.6a  
A:Map position: 5

Query Match 12.4%; Score 103; DB 2; Length 1558;  
Best Local Similarity 30.5%; Pred. No. 0.05;  
Matches 25; Conservative 7; Mismatches 26; Indels 24; Gaps 5;  
2Y 44 WKPGPWGRCTGCGPGVQSRVAVWC---DGWTSLSNCGESNRPPKERSC 93  
DB 405 WFTGDWESCSSTCGDQGGQYRVVYCHQVFANGRRVTEVDG-----NC-TVERPPVKQTC 457  
2Y 94 FR-VCDWHSDLFOWEVSQVHCHVLPYARGEVKPTAEVTAQGLQHR 114  
DB 458 NRFACP-----EWQAGPWSAC 473

## RESULT 8

134395  
hypothetical protein C37C3.6b - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 01-Dec-2000  
C:Accession: T34395; T34394  
R:Poustka, C.; Bradshaw, H.  
submitted to the EMBL Data Library, July 1996  
A:Description: The sequence of C. elegans cosmid C37C3.  
A:Reference number: Z21518  
A:Accession: T34395  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-2167 <GEI>  
A:Cross-references: EMBL:U64857; PIDN:AAC25868.1; GSPDB:GN00023; CESP:C37C3.6b  
A:Experimental source: strain Bristol N2; clone C37C3  
A:Accession: T34394  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-1555, 'SKP' <GE2>  
A:Cross-references: EMBL:U64857; PIDN:AAC25867.1; GSPDB:GN00023; CESP:C37C3.6a  
A:Experimental source: strain Bristol N2; clone C37C3

C:Genetics:  
A:Gene: CESP:C37C3.6b; CESP:C37C3.6a  
A:Map position: 5  
A:Introns: 32/3; 104/2; 156/2; 207/1; 459/2; 536/3; 577/2; 1105/3; 1367/1; 1438/1; 1556/

Query Match 12.4%; Score 103; DB 2; Length 2167;  
Best Local Similarity 30.5%; Pred. No. 0.067;  
Matches 25; Conservative 7; Mismatches 26; Indels 24; Gaps 5;  
QY 44 WKPGPWGRCTGCGPGVQSRVAVWC---DGWTSLSNCGESNRPPKERSC 93  
DB 405 WFTGDWESCSSTCGDQGGQYRVVYCHQVFANGRRVTEVDG-----NC-TVERPPVKQTC 457  
QY 94 FR-VCDWHSDLFOWEVSQVHCHVLPYARGEVKPTAEVTAQGLQHR 114  
DB 458 NRFACP-----EWQAGPWSAC 473

## RESULT 9

T47158  
hypothetical protein DKFZp762C110.1 - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000  
C:Accession: T47158  
R:Blum, H.; Bauersachs, S.; Mewes, H.W.; Weil, B.; Wiemann, S.  
submitted to the Protein Sequence Database, March 2000  
A:Reference number: Z24379  
A:Accession: T47158  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-550 <AAA>  
A:Cross-references: EMBL:AL162080  
A:Experimental source: adult melanoma (MeWo cell line); clone DKFZp762C110  
C:Genetics:  
A:Note: DKFZp762C110.1

Query Match 11.1%; Score 92; DB 2; Length 550;  
Best Local Similarity 27.4%; Pred. No. 0.24;  
Matches 31; Conservative 14; Mismatches 48; Indels 20; Gaps 8;  
QY 37 KKNQF-----TWKPGPWGRCTGCGPGVQSRVAVWC---DGWTSLSNCGESNRPPK 89  
DB 427 KKESFNAITFTSAWVTEWEGCSKSELGW-QRRVLECRDINGQPA--SECAKEVKPAS 483  
QY 90 ERSCEF-RVCDWHSDLFOWEVSQVHCHVLPYARGEVKPTAEVTAQGLQHR 140  
DB 484 TRPCADHPC-----QWQGEWSSCSKT-CGKG-YKRSKLKLSHDGGVLSH 528

## RESULT 10

T34565  
hypothetical protein DKFZp434K1772.1 - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T34565  
R:Poustka, A.; Wellenreuther, R.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, October 1999  
A:Reference number: Z21540  
A:Accession: T34565  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-111 <POU>  
A:Cross-references: EMBL:AL122085  
A:Experimental source: adult testis; clone DKFZp434K1772  
C:Genetics:  
A:Note: DKFZp434K1772.1

Query Match 10.9%; Score 90.5; DB 2; Length 111;  
Best Local Similarity 35.6%; Pred. No. 0.08;  
Matches 32; Conservative 8; Mismatches 43; Indels 7; Gaps 6;  
QY 44 WKPGPWGRCTGCGPGVQSRVAVWC---DGWTSLSNCGESNRPPKERSC-FRVCWDHSD 102

```
Db 11 WFTSPWSPGSRSC-QGGTQTRVQCLSTNQTLS--TRCPPQLRPSKRKPCNSQPCSORPD 67
QY 103 LFQWEVSDWHCVLPVYARGEVXP-RTAEC 131
Db 68 -DQCKDSS-PHCPLVQVQARLCVPPYTATC 95

RESULT 11
T09484
cartilage intermediate layer protein precursor - human
C:Species: Homo sapiens (man)
C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C:Accession: T09484
R:Lorenzo, P.; Neame, P.; Sommarin, Y.; Heinigard, D.
J. Biol. Chem. 273, 23469-23475, 1998
A:Title: Cloning and deduced amino acid sequence of a novel cartilage protein (CILP) ide
A:Reference number: 216689; MUID:98389785; PMID:9722584
A:Accession: T09484
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1184 <LOR>
A:Cross-references: EMBL:AF035408; NID:g3513502; PIDN:AAC33838.1; PID:g3513503
A:Experimental source: tissue type articular cartilage
C:Genetics:
A:Note: CILP
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-1184/Product: cartilage intermediate layer protein #status predicted <NAT>

Query Match 10.6%; Score 88; DB 2; Length 1184;
Best Local Similarity 34.0%; Pred. No. 1.1;
Matches 17; Conservative 9; Mismatches 18; Indels 6; Gaps 2;

QY 37 KKNQFTWKP-GPWGRTGDCGPGGVQSVRAVWCPHVDGWTSHLSNGESN 85
Db 145 RRDTERIWSPPWSPKSGACGGTGTTRICL-----AEMVSLCSEAS 189

RESULT 12
JT0573
retinoic acid-induced heparin-binding protein precursor - chicken
N:Alternate names: RIHB protein
C:Species: Gallus gallus (chicken)
C>Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 16-Jul-1999
C:Accession: JT0573; PT0032; S48199; S39266
R:Urios, P.; Duprez, D.; Le Caer, J.P.; Courtois, Y.; Vigny, M.; Laurent, M.
Biochem. Biophys. Res. Commun. 175, 617-624, 1991
A:Title: Molecular cloning of Ri-HB, a heparin binding protein regulated by retinoic aci
A:Reference number: JT0573; MUID:91207359; PMID:2018506
A:Accession: JT0573
A:Molecule type: mRNA
A:Residues: 1-142 <URI>
A:Experimental source: embryo
R:Raulais, D.; Lagente-Chevallier, O.; Guettet, C.; Duprez, D.; Courtois, Y.; Vigny, M.
Biochem. Biophys. Res. Commun. 174, 708-715, 1991
A:Title: A new heparin binding protein regulated by retinoic acid from chick embryo.
A:Reference number: PT0032; MUID:91128406; PMID:1993066
A:Accession: PT0032
A:Molecule type: protein
A:Residues: 22-77 <RAU>
R:Duprez, D.; Tregger, J.; Pecqueur, C.; Vigny, M.
Eur. J. Biochem. 224, 931-941, 1994
A:Title: Organization and promoter activity of the retinoic-acid-induced-heparin-binding
A:Reference number: S48199; MUID:95010085; PMID:7925417
A:Accession: S48199
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-88, 'R', 90-142 <DUP>
A:Cross-references: EMBL:X76482
R:Vigny, M.R.
submitted to the EMBL Data Library, December 1993
A:Description: Organization and promoter activity of the RIHB gene.
A:Reference number: S39266
A:Accession: S39266
```

```
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-88, 'R', 94-142 <DU2>
A:Cross-references: EMBL:X76482; NID:g434357; PIDN:CAA54020.1; PID:g434358
C:Genetics:
A:Introns: 27/1; 81/1; 135/1
C:Superfamily: pleiotrophin
C:Keywords: growth factor; heparin binding
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-142/Product: retinoic acid induced heparin-binding protein #status experimental <MA

Query Match 10.2%; Score 85; DB 2; Length 142;
Best Local Similarity 27.4%; Pred. No. 0.34;
Matches 37; Conservative 15; Mismatches 39; Indels 44; Gaps 8;

QY 18 RKLFLLSLLSHAHLKGGKDNQFIKPGP-----WGRC---TGDCGPGGVQSVRAVWC 68
Db 4 RGLLLLLLLLLAAAAAANKAKKMKKKGSECOQDWHGPGCIPIKSKOGLG----- 54
QY 69 FHVDGWTSHLSNGESNRPPKERSCFRVCWDHSLD---FQWEVSDWHCVLPVYARGEVX 125
Db 55 -YREG-----SCGDESKLK---CKIPCNWKKKFGADCKYKPFESWGGC---SAKTGVK 100
QY 126 PRT-----AEC 131
Db 101 TRSGILKKALYNAEC 115

RESULT 13
T18856
angiogenesis inhibitor homolog - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T18856; T24653
R:McMurray, A.
submitted to the EMBL Data Library, July 1995
A:Reference number: Z19031
A:Accession: T18856
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1444 <WIL>
A:Cross-references: EMBL:Z50004; PIDN:CAA90293.1; GSPDB:GN00028; CESP:CO2B4.1
A:Experimental source: clone CO2B4
R:McMurray, A.
submitted to the EMBL Data Library, July 1995
A:Reference number: Z19917
A:Accession: T24653
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1444 <WI2>
A:Cross-references: EMBL:Z50006; PIDN:CAA90302.1; GSPDB:GN00028; CESP:CO2B4.1
A:Experimental source: clone T07C5
C:Genetics:
A:Gene: CESP:CO2B4.1
A:Map position: X
A:Introns: 25/3; 70/3; 96/3; 139/3; 187/1; 234/2; 282/3; 376/2; 422/2; 478/3; 509/3; 561

Query Match 9.9%; Score 82; DB 2; Length 1444;
Best Local Similarity 21.3%; Pred. No. 5.3;
Matches 32; Conservative 16; Mismatches 48; Indels 54; Gaps 7;

QY 42 FIWKPGPGRCTGDCGPGGVQSVRAVWCPHVDGWTSHLSNGESNRPPKERSCFRVCWDHWS 101
Db 1069 FIW--SDWSSCSKSGQDGIQTRQKLCUFNN---AECSSVAESRRCKDLPSCSISGRT 1123
QY 102 -----DLFQW-EVSDWHHC-----VJ--VPVARGEVX 125
Db 1124 TSENGFDAPRWSEWSWSACSFCSTLTSTRRFPQVVDPTVQGFCAIGALEIQIPCAPGSCS 1183
QY 126 PR-----TAEQVTAQHGLOHRM 142
Db 1184 PSAGGWSLWSEWSKCKDGGTGHQIRNM 1213
```

Db 736 CEWSPCSTQLACEVGVQSRQCV 759

Search completed: February 24, 2004, 01:08:27  
Job time : 6.94835 secs

RESULT 14

I49607  
procollagen type V alpha 2 - mouse  
Species: Mus musculus (house mouse)  
Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 13-Aug-1999  
Accession: I49607  
Andrikopoulos, K.; Suzuki, H.R.; Solursh, M.; Ramirez, F.  
Dev. Dyn. 195, 113-120, 1992  
Title: Localization of pro-alpha 2(V) collagen transcripts in the tissues of the devel  
Reference number: I49607; PMID:1297453  
Accession: I49607  
Status: preliminary; translated from GB/EMBL/DDBJ  
Molecule type: mRNA  
Residues: 1-1497 <RES>  
Cross-references: GB:I02918; NID:G309180; PIDN:AAA37440.1; PID:G309181  
Genetics:  
Gene: Col5a-2  
Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;  
39-98/Domain: von Willebrand factor type C repeat homology <WVC>  
1270-1497/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 9.8%; Score 81.5; DB 2; Length 1497;  
Best Local Similarity 26.2%; Pred. No. 6.1;  
Matches 27; Conservative 12; Mismatches 33; Indels 31; Gaps 5;

2Y 14 WRSWRKFLILLSLLSHAAHLEGGKD-----NOFIWKGPGWGRCTG 54  
Db 5 WVGARFL-LILSVLLGYCVSIKAOENDEYDEIACIQHGQMYLNRIWKPSPCQICVC 63  
2Y 55 DCGPGGVQSRVAVCFVHDGWTSHLSNCGESNRPPKERSCFRVC 97  
Db 64 DNG-----AILCDKIE--CPEVLNCAPIPPGE--CCPVC 95

RESULT 15

I25061  
hypothetical protein T21B6.3 - Caenorhabditis elegans  
Species: Caenorhabditis elegans  
Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
Accession: T25061  
Cottage, A.  
submitted to the EMBL Data Library, November 1995  
Reference number: Z19975  
Accession: T25061  
Status: preliminary; translated from GB/EMBL/DDBJ  
Molecule type: DNA  
Residues: 1-788 <NIL>  
Cross-references: EMBL:Z68011; PIDN:CAA92014.1; GSPDB:GN00028; CESP:T21B6.3  
Experimental source: clone T21B6  
Genetics:  
Gene: CESP:T21B6.3  
Map position: X  
Introns: 20/1; 47/1; 76/1; 152/1; 735/2; 754/2

Query Match 9.6%; Score 79.5; DB 2; Length 788;  
Best Local Similarity 18.1%; Pred. No. 5.4;  
Matches 37; Conservative 18; Mismatches 62; Indels 87; Gaps 9;

2Y 8 TATCWWRSMRKFLILLSLLSHAAHLEGGKNO-----FIWKP-GPWGRCTGDCGP 58  
Db 564 SVTCGGGVAVRQRTCLGGVFGDHLG--QGPKTEQRACDGGPCSLSPWQEWSTCSACGS 621  
2Y 59 GGVSRAV-----WCFHVDGWTSHLSNCGESNRPPK 91  
Db 622 GMKRRQRCVQFGTDCQGPNEESFCYGPACAEWTEWC-----EWSGSSKCGPGQR-TRTR 676  
2Y 92 SCF-----RVCDSHSDLPQWEVSD----- 110  
Db 677 GCLGPNQBAATTCQGPSIETTLCEGSCCNW-SEWCHWSMCDKCCGGGVRYIEYNFTG 735  
2Y 111 --WHHCVLVPYARGEVKPRTAECV 132

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

» protein - protein search, using sw model

run on: February 23, 2004, 20:35:55 ; Search time 3.97635 Seconds  
(without alignments)  
1859.483 Million cell updates/sec

Title: US-10-022-710-6\_COPY\_1\_142

Perfect score: 830

Sequence: 1 MFPKSNLVTQVWWSMRKL.....EVKPTAEVTAQHGLOHRM 142

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	143	17.2	1906	1	AT20 MOUSE
2	128.5	15.5	1095	1	AT17 HUMAN
3	127	15.3	1935	1	AT19 HUMAN
4	123.5	14.9	1210	1	AT19 MOUSE
5	121	14.6	1911	1	AT20 HUMAN
6	116	14.0	562	1	AT15 MOUSE
7	114	13.7	968	1	AT15 MOUSE
8	113.5	13.7	1213	1	AT15 MOUSE
9	112.5	13.6	1223	1	AT14 HUMAN
10	111	13.4	950	1	AT15 HUMAN
11	110.5	13.3	1207	1	AT19 HUMAN
12	110	13.3	967	1	AT19 RAT
13	109.5	13.2	1211	1	AT19 HUMAN
14	108.5	13.1	1077	1	AT10 HUMAN
15	107.5	13.0	1205	1	AT10 BOVIN
16	103	12.4	525	1	AT11 HUMAN
17	102.5	12.3	1205	1	AT13 HUMAN
18	102	12.3	1593	1	AT12 HUMAN
19	100.5	12.1	1224	1	AT16 HUMAN
20	99.5	12.0	377	1	AT11 MOUSE
21	96.5	11.6	1081	1	AT18 HUMAN
22	96	11.6	890	1	AT18 HUMAN
23	92	11.1	967	1	AT18 HUMAN
24	88.5	10.7	323	1	TNR6 BOVIN
25	88.5	10.7	450	1	AT10 MOUSE
26	85.5	10.3	930	1	AT15 MOUSE
27	85	10.2	142	1	MK_CHICK
28	81.5	9.8	630	1	AT14 RAT
29	81.5	9.8	930	1	AT15 HUMAN
30	79.5	9.6	1584	1	BAL1 HUMAN
31	77.5	9.3	860	1	AT15 HUMAN
32	77	9.3	905	1	AT15 MOUSE
33	76.5	9.2	143	1	MK_HUMAN

RESULT 1  
AT20\_MOUSE  
ID AT20\_MOUSE STANDARD; PRT; 1906 AA.  
AC P59511;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE ADAMTS-20 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 20) (ADAM-TS 20) (ADAM-TS20).  
GN ADAMTS20.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCB1\_TaxID=10090;  
[1]  
SEQUENCE FROM N.A. (ISOFORM 1).  
RC TISSUE=Fetal brain;  
RX MEDLINE=22566039; PubMed=12562771;  
RA Llamazares M., Cal S., Quesada V., Lopez-Otin C.; Identification and characterization of ADAMTS-20 defines a novel subfamily of metalloproteinases-disintegrins with multiple thrombospondin-1 repeats and a unique GON domain.";  
RT J. Biol. Chem. 278:13382-13389(2003).  
[2]  
SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND DISEASE.  
RA MEDLINE=22806432; PubMed=12925592;  
RA Rao C., Feenizler D., Loftus S.K., Liu S., McPherson J.D., Jungers K.A., Apte S.S., Pavan W.J., Beier D.R.; "A defect in a novel ADAMTS family member is the cause of the belted white-spotting mutation.";  
RT Development 130:4665-4672(2003).  
CC -!- FUNCTION: May play a role in tissue-remodeling process occurring in both normal and pathological conditions.  
CC -!- COPACFOR: Binds 1 zinc ion per subunit (By similarity).  
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (By similarity).  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=1; Synonyms=ADAMTS20 B long isoform;  
CC IsoId=P59511-1; Sequence=displayed;  
CC Name=2; Synonyms=ADAMTS20 A short isoform;  
CC IsoId=P59511-2; Sequence=VSP\_007606, VSP\_007607;  
CC -!- TISSUE SPECIFICITY: Expressed at low level in testis and brain.  
CC -!- PTM: The precursor is cleaved by a furin endopeptidase (By similarity).  
CC -!- DISEASE: Defects in ADAMTS20 are the cause of the belted (bt) phenotype. It is a pigmentary defect which occurs as a result of a defect in melanocyte development.  
CC -!- SIMILARITY: Belongs to peptidase family M12B.  
CC -!- SIMILARITY: Contains 1 disintegrin-like domain.  
CC -!- SIMILARITY: Contains 1 GON domain.  
CC -!- SIMILARITY: Contains 15 TSP type-1 domains.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration

075173 homo sapien  
062217 mus musculus  
077736 sus scrofa  
P05997 homo sapien  
Q13591 homo sapien  
Q9WJ19 mus musculus  
Q04592 mus musculus  
P11218 urtica dioi  
Q95116 bos taurus  
Q60242 homo sapien  
P14728 xanthomonas

## ALIGNMENTS

34 76.5 9.2 837 1 AT54 HUMAN  
35 76.5 9.2 1077 1 TNR6 MOUSE  
36 76 9.2 332 1 TNR6 PIG  
37 74 8.9 1496 1 CA25 HUMAN  
38 73.5 8.9 1074 1 SM5A HUMAN  
39 72.5 8.7 4543 1 FBL4 MOUSE  
40 72.5 8.7 4544 1 LRPI HUMAN  
41 71.5 8.6 1877 1 PK55 MOUSE  
42 71 8.6 372 1 AG1 URDI  
43 71 8.6 1170 1 TSP2 BOVIN  
44 71 8.6 1522 1 BAL1 HUMAN  
45 70 8.4 784 1 YAV2\_XANCV

between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; AJ512753; CAD54808.3; -  
EMBL; AY189815; AAO74895.1; -  
EMBL; AY189816; AAO74896.1; -  
MGP; MGI:2660628; Adams20.  
InterPro; IPR001762; Disintegrin.  
InterPro; IPR006025; Pept\_M\_Zn\_BS.  
InterPro; IPR001590; Peptidase\_M12B.  
InterPro; IPR002870; Peptidase\_M12B\_N.  
InterPro; IPR000884; TSP1.  
Pfam; PF01562; Pep\_M12B\_propep; 1.  
Pfam; PF01421; Reptolysin; 1.  
Pfam; PF00090; TSP; 1.  
SMART; SM00209; TSP; 14.  
PROSITE; PS00215; ADAM\_MEPRO; 1.  
PROSITE; PS00427; DISINTEGRIN\_1; FALSE\_NEG.  
PROSITE; PS00214; DISINTEGRIN\_2; FALSE\_NEG.  
PROSITE; PS0092; TSP1; 13.  
PROSITE; PS00142; ZINC\_PROTEASE; 1.  
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;  
Repeat; Extracellular matrix; Alternative splicing.  
FT SIGNAL 1 26  
FT PROPEP 27 249  
FT CHAIN 250 1906  
FT DOMAIN 250 464  
FT DOMAIN 465 552  
FT DOMAIN 553 608  
FT DOMAIN 609 720  
FT DOMAIN 721 842  
FT DOMAIN 843 901  
FT DOMAIN 906 962  
FT DOMAIN 962 1015  
FT DOMAIN 1017 1074  
FT DOMAIN 1075 1131  
FT DOMAIN 1148 1202  
FT DOMAIN 1203 1260  
FT DOMAIN 1300 1351  
FT DOMAIN 1354 1411  
FT DOMAIN 1412 1465  
FT DOMAIN 1468 1526  
FT DOMAIN 1527 1584  
FT DOMAIN 1585 1648  
FT DOMAIN 1650 1706  
FT DOMAIN 1707 1906  
FT METAL 399 399  
FT ACT\_SITE 400 400  
FT METAL 403 403  
FT METAL 403 409  
FT CARBOHYD 92 92  
FT CARBOHYD 221 221  
FT CARBOHYD 714 714  
FT CARBOHYD 798 798  
FT CARBOHYD 805 805  
FT CARBOHYD 1057 1057  
FT CARBOHYD 1562 1562  
FT CARBOHYD 1719 1719  
FT CARBOHYD 1759 1759  
FT CARBOHYD 1777 1777  
FT VARSPPLIC 1424 1425  
FT Missing (in isoform 2).  
FT VARSPPLIC 1426 1906  
FT CONFLICT 1211 1211  
FT CONFLICT 1262 1262  
FT SEQUENCE 1906 AA; 212041 MW; 0EB2568547B557EA CRC64;  
Query Match 17.2%; Score 143; DB 1; Length 1906;

Best Local Similarity 32.4%; Pred. No. 1.4e-06;  
Matches 35; Conservative 14; Mismatches 41; Indels 18; Gaps 6;  
QY 38 KXNQIPWPGPWRGTGCGGGVQSVRAVCFHVDGWTSHLSNCGESNRPKERSCFRVC 97  
DB 1300 RGQV--WRTGPGAGCSRSCA-GGLQHRVAVCODEDGRSA--TSCGGSKPPESRHC----- 1350  
QY 98 DWHSDFQWEVSDWHCHVLPYARGVEKPRTAEC-----VTAQHGLQ 139  
DB 1351 -GSGPCPHNYGDWGECTQT--CGGVKSRFVICQPPNGQMTQEHSC 1395  
RESULT 2  
AT17\_HUMAN  
ID AT17\_HUMAN STANDARD; PRT; 1095 AA.  
AC Q8TE56;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE ADAMTS-17 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase  
DE with thrombospondin motifs 17) (ADAM-TS 17) (ADAM-TS17).  
GN ADAMTS17.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21856482; PubMed=11867212;  
RA Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,  
RA Lopez-Otin C.;  
RT "Cloning, expression analysis, and structural characterization of  
RT seven novel human ADAMTSs, a family of metalloproteinases with  
RT disintegrin and thrombospondin-1 domains.";  
RL Gene 283:49-62(2002).  
CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular  
CC matrix (By similarity).  
CC -1- TISSUE SPECIFICITY: Expressed in fetal lung, in adult brain,  
CC prostate, and liver.  
CC -1- PTM: The precursor is cleaved by a furin endopeptidase (By  
CC similarity).  
CC -1- SIMILARITY: Belongs to peptidase family M12B.  
CC -1- SIMILARITY: Contains 1 disintegrin-like domain.  
CC -1- SIMILARITY: Contains 1 PLAC domain.  
CC -1- SIMILARITY: Contains 5 TSP type-1 domains.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).



Somerville R.P., Longpre J.-M., Jungers K.A., Engle J.M., Ross M.,  
 Ravankar S., Wight T.N., Leduc R., Apte S.S.;  
 "Characterization of ADAMTS-9 and ADAMTS-20 as a distinct ADAMTS  
 subfamily related to Caenorhabditis elegans GON-1."; *J. Biol. Chem.* 278:9503-9513(2003).  
 [3]  
 SEQUENCE OF 159-1935 FROM N.A. (ISOFORM 2).  
 TISSUE=Brain;  
 MEDLINE=20181126; PubMed=10718198;  
 Nagase T., Kikuno R., Ishikawa K.-I., Hirotsawa M., Ohara O.;  
 "Prediction of the coding sequences of unidentified human genes. XVI.  
 The complete sequences of 150 new cDNA clones from brain which code  
 for large proteins in vitro."; *DNA Res.* 7:165-73(2000).  
 CC -!- FUNCTION: Cleaves the large aggregating proteoglycans, aggrecan  
 and versican.  
 CC -!- CATALYTIC ACTIVITY: Cleaves aggrecan at the 1838-Glu-|-Ala-1839  
 site and versican at the 1428-Glu-|-Ala-1429 site.  
 CC -!- COFACTOR: Binds 1 zinc ion per subunit [By similarity].  
 CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular  
 matrix [By similarity].  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=3;  
 CC Name=1; Synonyms=ADAMTS-9B;  
 CC IsoId=Q9P2N4-3; Sequence=Displayed;  
 CC Name=2; Synonyms=Long;  
 CC IsoId=Q9P2N4-1; Sequence=VSP\_007548; VSP\_007549;  
 CC Note=Way result from the retention of an intron in the cDNA  
 leading to a premature stop codon;  
 CC Name=3; Synonyms=Short;  
 CC IsoId=Q9P2N4-2; Sequence=VSP\_005499; VSP\_005500;  
 CC TISSUE SPECIFICITY: Highly expressed in all fetal tissues  
 CC Expressed in a number of adult tissues with highest expression in  
 CC heart, placenta and skeletal muscle.  
 CC -!- DOMAIN: The spacer domain and the TSP type-1 domains are important  
 CC for a tight interaction with the extracellular matrix [By  
 CC similarity].  
 CC -!- DOMAIN: The ancillary domains, including the TSPs domain, are  
 CC required for specific extracellular localization and for its  
 CC versicanase and aggrecanase activities.  
 CC -!- PTM: The precursor is cleaved by a furin endopeptidase [By  
 CC similarity].  
 CC -!- SIMILARITY: Belongs to peptidase family M12B.  
 CC -!- SIMILARITY: Contains 1 disintegrin-like domain.  
 CC -!- SIMILARITY: Contains 1 GON domain.  
 CC -!- SIMILARITY: Contains 15 TSP type-1 domains.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AF261918; AAF89106.1; -;  
 CC EMBL; AF488803; AAO:5765.1; -;  
 CC EMBL; AB037733; BAA92550.1; -;  
 CC HSSP; P15167; 1ATL.  
 CC MEROPS; M12\_021; -;  
 CC GenSeq; HGNC:13202; ADAMTS9.  
 CC MIM; 605421; -;  
 CC GO; GO:000823; F:metallopeptidase activity; TAS.  
 CC GO; GO:000725; P:development; TAS.  
 CC GO; GO:0006516; P:glycoprotein catabolism; TAS.  
 CC InterPro; IPR001762; Disintegrin.  
 CC InterPro; IPR001818; Pept\_M10A\_M12B.  
 CC InterPro; IPR006025; Pept\_M\_Zn\_BS.  
 CC InterPro; IPR001590; Peptidase\_M12B.  
 CC InterPro; IPR002870; Peptidase\_M12B\_N.  
 CC InterPro; IPR000894; TSP1.  
 CC Pfam; PF01562; Pp\_M12B\_propep; 1.  
 CC Pfam; PF01421; Repolysin; 1.

DR PFAM; PF00090; tsp.1; 11.  
 DR SMART; SM00209; TSPI; 12.  
 DR PROSITE; PS00215; ADAM\_MEPRO; 1.  
 DR PROSITE; PS00546; CYSTEINE\_SWITCH; FALSE\_NEG.  
 DR PROSITE; PS00427; DISINTEGRIN\_1; FALSE\_NEG.  
 DR PROSITE; PS00214; DISINTEGRIN\_2; FALSE\_NEG.  
 DR PROSITE; PS00092; TSPI; 14.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 DR Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;  
 KW Repeat; Extracellular matrix; Alternative splicing.  
 FT SIGNAL 1 18  
 FT PROPEP 19 287  
 FT CHAIN 288 1935  
 FT DOMAIN 293 499  
 FT DOMAIN 509 587  
 FT DOMAIN 588 643  
 FT DOMAIN 644 752  
 FT DOMAIN 753 877  
 FT DOMAIN 878 936  
 FT DOMAIN 939 997  
 FT DOMAIN 998 1049  
 FT DOMAIN 1052 1109  
 FT DOMAIN 1110 1166  
 FT DOMAIN 1182 1240  
 FT DOMAIN 1241 1296  
 FT DOMAIN 1328 1379  
 FT DOMAIN 1382 1440  
 FT DOMAIN 1441 1494  
 FT DOMAIN 1497 1555  
 FT DOMAIN 1556 1611  
 FT DOMAIN 1612 1676  
 FT DOMAIN 1677 1734  
 FT DOMAIN 1735 1935  
 FT DOMAIN 88 96  
 FT SITE 223 223  
 FT METAL 434 434  
 FT ACT\_SITE 435 435  
 FT METAL 438 438  
 FT METAL 444 444  
 FT CARBOHYD 112 112  
 FT CARBOHYD 135 135  
 FT CARBOHYD 271 271  
 FT CARBOHYD 749 749  
 FT CARBOHYD 840 840  
 FT CARBOHYD 1213 1213  
 FT CARBOHYD 1267 1267  
 FT CARBOHYD 1788 1788  
 FT CARBOHYD 1806 1806  
 FT VARSPLIC 1064 1072  
 FT VARSPLIC 1073 1935  
 FT VARSPLIC 1624 1629  
 FT VARSPLIC 1630 1935  
 FT CONFLICT 46 46  
 FT CONFLICT 96 96  
 FT CONFLICT 182 182  
 FT CONFLICT 367 367  
 FT CONFLICT 1117 1117  
 FT SEQUENCE 1935 AA; 216556 MW; FD3D51E8830A3C6 CRC64;  
 Query Match 15.3%; Score 127; DB 1; Length 1935;  
 Best Local Similarity 31.5%; Pred. No. 6e-05;  
 Matches 28; Conservative 15; Mismatches 34; Indels 12; Gaps 5;  
 2y 44 WKGPWGRCTGDCGPGVQSRVAVCFPHDWTSHLNCGESNPPKERC-PRVCDWHS 102  
 1332 WRTGPGWACSSSTCA-GGSRVRVVCDDNGYTA--NDCVERIKPDQRACSGPCP---- 1384  
 2y 103 LFQWESDWHCVLPVYARGEKPRTAEC 131  
 Db 1385 --QWAYGNWGECTKL--CGGGIRTRLVVC 1409  
 RESULT 4  
 AT19 MOUSE  
 ID AT19 MOUSE STANDARD; PRT; 1210 AA.  
 AC P59509;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE ADAMTS-19 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase  
 DE with thrombospondin motifs 19) (ADAM-TS 19) (ADAM-TS19).  
 GN ADAMTS19.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Ovary;  
 RX MEDLINE=22505168; PubMed=12617826;  
 RA Menke D.B., Page D.C.;  
 RT "Sexually dimorphic gene expression in the developing mouse gonad.";  
 RL Gene Expr. Patterns 2:359-367(2002).  
 CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular  
 CC matrix (By similarity).  
 CC -!- TISSUE SPECIFICITY: Expressed predominantly in fetal ovary, low  
 CC levels of expression is also detected in kidney, heart, skeletal  
 CC muscle, lung and testis.  
 CC -!- DEVELOPMENTAL STAGE: Expression is strongest in anterior and  
 CC ventral regions of the ovary at 12.5 and 13.5 dpc before becoming  
 CC more uniform.  
 CC -!- PTM: The precursor is cleaved by a furin endopeptidase (By  
 CC similarity).  
 CC -!- SIMILARITY: Belongs to peptidase family M12B.  
 CC -!- SIMILARITY: Contains 1 disintegrin-like domain.  
 CC -!- SIMILARITY: Contains 5 TSP type-1 domains.  
 CC -!- SIMILARITY: Contains 1 PLAC domain.  
 CC -!- CAUTION: By homology with the human sequence, it is uncertain  
 CC whether Met-1 or Met-5 is the initiator.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 CC EMBL; AY135183; MAM10155.1; -  
 CC MGD; MGI:2442875; Adamts19.  
 CC InterPro; IPR006586; ADAM\_cysteine.  
 CC InterPro; IPR001762; Disintegrin.  
 CC InterPro; IPR001818; Pept\_M10A\_M12B.  
 CC InterPro; IPR006025; Pept\_M\_Zn\_BS.  
 CC InterPro; IPR001590; Peptidase\_M12B.  
 CC InterPro; IPR003870; Peptidase\_M12B\_N.  
 CC InterPro; IPR000884; TSPI.  
 CC Pfam; PF01562; Pep\_M12B\_propep; 1.  
 CC Pfam; PF01421; Repolysin; 1.  
 CC Pfam; PF00090; tsp.1; 5.  
 CC SMART; SM00608; ACR; 1.  
 CC SMART; SM00209; TSPI; 5.  
 CC PROSITE; PS02015; ADAM\_MEPRO; 1.  
 CC PROSITE; PS00546; CYSTEINE\_SWITCH; FALSE\_NEG.  
 CC PROSITE; PS00427; DISINTEGRIN\_1; FALSE\_NEG.  
 CC PROSITE; PS02014; DISINTEGRIN\_2; FALSE\_NEG.  
 CC PROSITE; PS00300; PLAC; 1.  
 CC PROSITE; PS00092; TSPI; 5.  
 CC PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 CC Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;  
 KW Repeat; Extracellular matrix.

```

FT SIGNAL 1 30
FT PROPEP 31 319
FT CHAIN 320 1210
FT DOMAIN 320 548
FT DOMAIN 549 636
FT DOMAIN 637 689
FT DOMAIN 690 783
FT DOMAIN 794 917
FT DOMAIN 918 978
FT DOMAIN 979 1040
FT DOMAIN 1042 1086
FT DOMAIN 1090 1147
FT DOMAIN 1163 1202
FT SITE 297 297
FT METAL 485 485
FT ACT_SITE 486 486
FT METAL 489 489
FT METAL 495 495
FT CARBOHYD 54 54
FT CARBOHYD 263 263
FT CARBOHYD 800 800
FT CARBOHYD 910 910
FT CARBOHYD 931 931
FT CARBOHYD 952 952
FT CARBOHYD 1012 1012
SQ SEQUENCE 1210 AA; 134560 MW; 0AB812ABAB4BB7A2 CRC64;

Query Match 14.98; Score 123.5; DB 1; Length 1210;
Best Local Similarity 30.24; Pred. No. 8.5e-05;
Matches 32; Conservative 17; Mismatches 40; Indels 17; Gaps 7;

Qy 35 EGKDNDFIWKPGWGRCTGDCGGVQSVRAVWCFVHDGWTSHLSNCGSNRPKERSCF 94
Db 1035 EG-QDQMTVEAGWSVCSVKCGK-GVRHRTVC-----TNRKKCVLSTPREAED-- 1084

Qy 95 RVDWHSDLQFQWEVSDVHHCVLPYARGEVKPTAGC---VTAQHG 137
Db 1085 --CEDYSKYVVRVGDWSKC-SITCGKG-MQSRVIQMKHKITGRHG 1126

RESULT 5
AT20_HUMAN
ID AT20_HUMAN STANDARD; PRT; 1911 AA.
AC P59510;
DT 10-OCT-2003 (Rel. 42, Created)
DD 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE ADAMTS-20 precursor (EC 3.4.24.-) (A disintegrin and
DE metalloproteinase with thrombospondin motifs 20) (ADAM-TS 20) (ADAM-
DE TS20).
EN Homo sapiens (Human).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=22513925; PubMed=12514189;
RA Somerville R.P., Longpre J.-M., Jungers K.A., Engle J.M., Ross M.,
RA Evanko S., Wight T.N., Leduc R., Apté S.S.;
RA "Characterization of ADAMTS-9 and ADAMTS-20 as a distinct ADAMTS
RT subfamily related to Caenorhabditis elegans GON-1.";
RL J. Biol. Chem. 278:9503-9513(2003).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RX TISSUE=Liver;
RA MEDLINE=22566039; PubMed=12562771;
RA Llamazares M., Cal S., Quesada V., Lopez-Otin C.;
RT "Identification and characterization of ADAMTS-20 defines a novel
RT subfamily of metalloproteinases-disintegrins with multiple
RT thrombospondin-1 repeats and a unique GON domain.";
RL J. Biol. Chem. 278:13382-13389(2003).
CC -!- FUNCTION: May play a role in tissue-remodeling process occurring

in both normal and pathological conditions.
-!- COPACITOR: Binds 1 zinc ion per subunit (By similarity).
-!- SURCELLULAR LOCATION: Secreted. Associated with the extracellular
matrix (By similarity).
-!- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=2;
Name=1;
IsoId=P59510-1; Sequence=Displayed;
Name=2;
IsoId=P59510-2; Sequence=VSP 007106, VSP 007107, VSP 007108;
-!- TISSUE SPECIFICITY: Very sparingly expressed, although is detected
at low levels in testis, prostate, ovary, heart, placenta, lung
and pancreas. Overexpressed in several brain, colon and breast
carcinomas.
-!- PTM: The precursor is cleaved by a furin endopeptidase (By
similarity).
-!- SIMILARITY: Belongs to peptidase family M12B.
-!- SIMILARITY: Contains 1 disintegrin-like domain.
-!- SIMILARITY: Contains 1 GON domain.
-!- SIMILARITY: Contains 15 TSP type-1 domains.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
-----
EMBL; AF488804; AAO15766.1; -
EMBL; AJ515153; CAD56159.3; -
EMBL; AJ515154; CAD56160.2; -
Genew; HGNC:17178; ADAMTS20.
InterPro; IPR001762; Disintegrin.
InterPro; IPR006025; Pept_M_Zn_BS.
InterPro; IPR001590; Peptidase_M12B.
InterPro; IPR002870; Peptidase_M12B_N.
InterPro; IPR000884; TSP1.
Pfam; PF01562; Pep_M12B_propep; 1.
Pfam; PF01421; Reprolysin; 1.
Pfam; PF00090; Tsp_1; 11.
SMART; SMO0209; TSP1; 12.
PROSITE; PS02015; ADAM_MEPRO; 1.
PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
PROSITE; PS02014; DISINTEGRIN_2; FALSE_NEG.
PROSITE; PS50092; TSP1; 12.
PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
Repeat; Extracellular matrix; Alternative splicing.
FT SIGNAL 1 21
FT PROPEP 22 253
FT CHAIN 254 1911
FT DOMAIN 254 467
FT DOMAIN 468 555
FT DOMAIN 556 611
FT DOMAIN 612 723
FT DOMAIN 724 846
FT DOMAIN 847 905
FT DOMAIN 906 962
FT DOMAIN 967 1024
FT DOMAIN 1025 1074
FT DOMAIN 1077 1136
FT DOMAIN 1153 1207
FT DOMAIN 1208 1265
FT DOMAIN 1305 1357
FT DOMAIN 1359 1417
FT DOMAIN 1418 1476
FT DOMAIN 1477 1532
FT DOMAIN 1536 1589
FT DOMAIN 1590 1653
FT DOMAIN 1655 1711
FT DOMAIN 1712 1911
FT METAL 403 403
ZINC (CATALYTIC) (BY SIMILARITY).

```



```
SQ SEQUENCE 562 AA; 60963 MW; 48009AD337D4E4A CRC64;
Query Match 14.0% Score 116; DB 1; Length 562;
Best Local Similarity 35.1%; Pred.No. 0.00023;
Matches 33; Conservative 9; Mismatches 40; Indels 12; Gaps 6;
2Y 44 WKGPWRCCTGDCGGVQSRVWCVHVDGWTSHLNCGSENNPPKXSCFRVCDWHSDL 103
Db 455 WVAGSWGFCVSCG-SGLQKRAVDRCRSPG-QQGASACVDHPRLEKRCAGCEPCP----- 507
2Y 104 FQEVSDPHHCVLVPYARGVKPRTAEVCYTAQHG 137
Db 508 -TWELGNWSPC-SKSCORG-FKRPLKCV--GGH 536

RESULT 7
ATSI_MOUSE
ID -ATSI_MOUSE STANDARD; PRT; 968 AA.
AC P97857; 054768;
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE ADAMTS-1 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1).
EN ADAMTS1
OS Mus musculus (Mouse).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SvJ.
RX MEDLINE=98110593; PubMed=9441751;
XA Kuno K., Lizasa H., Ohno S., Matsushima K.;
RT "The exon/intron organization and chromosomal mapping of the mouse
RT ADAMTS-1 gene encoding an ADAM family protein with TSP motifs.";
RL J. Biol. Chem. 272:556-582(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97150761; PubMed=8995297;
XA Kuno K., Kanada N., Nakashima E., Fujiki F., Ichimura F.,
RA Matsushima K.;
RT "Molecular cloning of a gene encoding a new type of metalloproteinase-
RT disintegrin family protein with thrombospondin motifs as an
RT inflammation associated gene.";
RL J. Biol. Chem. 272:556-582(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Limb, and Mammary gland;
RX MEDLINE=22388257; PubMed=12477932;
XA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
XA Klausner R.D., Collins F.S., Wagner L., Shenmen C.W., Schuler G.D.,
XA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
XA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
XA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
XA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,
XA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prance C.,
XA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
XA Rosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
XA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
XA Villalón D.K., Muny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
XA Fahey J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,
XA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
XA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
XA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
XA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
XA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RP CHARACTERIZATION, AND MUTAGENESIS OF GIU-403.
RX MEDLINE=99303657; PubMed=10373500;
```

```
RA Kuno K., Terashima Y., Matsushima K.;
RT "ADAMTS-1 is an active metalloproteinase associated with the
RT extracellular matrix.";
RL J. Biol. Chem. 274:18821-18826(1999).
RN [5]
RP FUNCTION.
RX MEDLINE=20389568; PubMed=10930576;
RA Kuno K., Okada Y., Kawashima H., Nakamura H., Miyasaka M.,
RA Ohno H., Matsushima K.;
RT "ADAMTS-1 cleaves a cartilage proteoglycan, aggrecan.";
RL FEBS Lett. 478:241-245(2000).
RN [6]
RP FUNCTION, AND INDUCTION.
RX MEDLINE=20243757; PubMed=10781075;
RA Robker R.L., Russell D.L., Espey L.L., Lydon J.P., O'Malley B.W.,
RA Richards J.S.;
RT "Progesterone-regulated genes in the ovulation process: ADAMTS-1 and
RT cathepsin L proteases.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:4689-4694(2000).
CC -!- FUNCTION: Cleaves aggrecan, a cartilage proteoglycan, and may be
CC involved in its turnover. Has angiogenic inhibitor activity (By
CC similarity). Active metalloproteinase, which may be associated with
CC various inflammatory processes as well as development of cancer
CC cachexia. May play a critical role in follicular rupture (By
CC similarity).
CC -!- CATALYTIC ACTIVITY: Cleaves aggrecan at the 1691-Glu-|-Leu-1692
CC site, within the chondroitin sulfate attachment domain.
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix.
CC -!- INDUCTION: Induced in vitro in colon adenocarcinoma cells by
CC interleukin-1, or in vivo in kidney and heart by
CC lipopolysaccharide. Also induced by LH stimulation in granulosa
CC cells of preovulatory follicles.
CC -!- DOMAIN: The spacer domain and the TSP type-1 domains are important
CC for a tight interaction with the extracellular matrix.
CC -!- PTM: The precursor is cleaved by a furin endopeptidase.
CC -!- SIMILARITY: Belongs to peptidase family M12B.
CC -!- SIMILARITY: Contains 1 disintegrin-like domain.
CC -!- SIMILARITY: Contains 3 TSP type-1 domains.
CC -!- CAUTION: Ref.2 sequence differs from that shown due to a
CC frameshift in position 7.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
EMBL; AB001735; BA24501.1; ALT_INIT.
EMBL; D67076; BA11088.1; ALT_FRAME.
EMBL; BC040382; AAH40382.1; -.
EMBL; BC050834; AAH50834.1; -.
MOP; M12.222; -.
MOP; MGI:109249; Adamts1.
InterPro; IPR006586; ADAM_cysteine.
InterPro; IPR001762; Disintegrin.
InterPro; IPR006025; Pept_M_12B.
InterPro; IPR001590; Peptidase_M12B.
InterPro; IPR002870; Peptidase_M12B_N.
InterPro; IPR000884; TSP1.
InterPro; IPR003085; TSP1.
Pfam; PF01562; Pep_M12B_propep; 1.
Pfam; PF01421; Repolysin; 1.
Pfam; PF00090; tsp_1; 3.
PRINTS; PR01705; TSP1REPEAT.
SMART; SMC0608; ACR; 1.
SMART; SMC0209; TSP1; 3.
PROSITE; PS50215; ADAM_MOPRO; 1.
PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
PROSITE; PS50214; DISINTEGRIN_2; FALSE_NEG.
```

DR PROSITE; PS50092; TSP1; 3.  
 DR PROSITE; PS00142; ZINC PROTEASE; 1.  
 KW Hydrolyase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;  
 KW Repeat; Extracellular matrix; Heparin-binding.  
 FT SIGNAL 1 48  
 FT PROPEP 49 253  
 FT CHAIN 254 968  
 FT DOMAIN 254 476  
 FT DOMAIN 477 559  
 FT DOMAIN 560 615  
 FT DOMAIN 616 725  
 FT DOMAIN 726 850  
 FT DOMAIN 851 911  
 FT DOMAIN 912 968  
 FT DOMAIN 195 159  
 FT SITE 206 206  
 FT METAL 402 402  
 FT ACT SITE 403 403  
 FT METAL 406 406  
 FT METAL 412 412  
 FT CARBOHYD 546 548  
 FT CARBOHYD 721 721  
 FT CARBOHYD 765 765  
 FT CARBOHYD 783 783  
 FT CARBOHYD 946 946  
 FT MUTAGEN 403 403  
 FT CONFLICT 335 335  
 FT CONFLICT 425 425  
 FT SEQUENCE 968 AA; 425841 MW; 42EBDA55499FB6C1 CRC64;  
 Query Match 13.7%; Score 114; DB 1; Length 968;  
 Best Local Similarity 28.7%; Pred. No. 0.00064;  
 Matches 37; Conservative 15; Mismatches 49; Indels 28; Gaps 8;  
 QY 26 LLLSH-----AHLEGGKQNF-----IWKEGPGWGRCTGCGPGVGSRAVWCVHV 71  
 DB 827 LMGVHALRPKIKFTFMKKKTSFNAIPTFSEWVEEWGEGSKTCG-SGWRVVVQCRDI 885  
 QY 72 DQWTHSLSCGNSPPKRSFCFRCDWHSDL--FQWESVDWHCVLPYPARGVEVKPETA 129  
 DB 886 NHFPA--SECAKEVFASTRPC-----ADLPCHQVQWDSPSKT-CGKG-YKGRKL 934  
 QY 130 ECVTAQHL 138  
 DB 935 KCVSHDGGV 943

## RESULT 8

AT52\_MOUSE  
 ID AT52\_MOUSE STANDARD; PRT; 1213 AA.  
 AC Q8C9W3;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE ADAMTS-2 precursor (SC 3.4.24.14) (A disintegrin and metalloproteinase  
 DE with thrombospondin motifs 2) (ADAM-TS2) (Procollagen I/II  
 DE amino-propeptide processing enzyme) (Procollagen I N-proteinase) (PC  
 DE I-NP) (Procollagen N-endopeptidase) (pNPI).  
 GN ADAMTS2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 CX NCBI\_TaxID=10090;  
 RN [1]  
 SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; Tissue=Thymus;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA Nikaide Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
 RA Okazaki I., Otsu N., Saito K., Suzuki H., Yamanaka I., Kiyosawa H.,  
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
 RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,

Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
 RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
 RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,  
 RA Konagaya A., Kurochkin I.V., Lee Y., Lennard B., Lyons P.A.,  
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,  
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
 RA Ravelin A., Schneider C., Sempile C.A., Setou M., Shimada K.,  
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
 RA Verardo R., Wagner L., Wohlested C., Wang Y., Watanabe Y., Wells C.,  
 RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
 RA Hiroxane-Kishikawa T., Kono H., Nakamura M., Sakazume N., Sato K.,  
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
 RA Birney E., Hayashizaki Y.;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs.";  
 RN Nature 420:563-573(2002).  
 [2]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN=FVB/N; Tissue=Breast tumor;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Krausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Sapotnick M., Soares M.B., Donald M.P., Casavant T.L., Schetz T.E.,  
 RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Bask S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huiyk S.W.,  
 RA Villalón D.K., Morley D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman A.C., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalish D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -!- FUNCTION: Cleaves the propeptides of type I and II collagen prior  
 CC to fibril assembly. Does not act on type III collagen. May also  
 CC play a role in development that is independent of its role in  
 CC collagen biosynthesis (By similarity).  
 CC -!- CATALYTIC ACTIVITY: Cleaves the N-propeptide of collagen chain  
 CC at alpha-1(I) at Pro-|-Gln and of alpha-1(II) and alpha-2(I) chains  
 CC at Ala-|-Gln.  
 CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).  
 CC -!- SUBUNIT: May belong to a multimeric complex. Binds specifically to  
 CC collagen type XIV (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular  
 CC matrix (By similarity).  
 CC -!- DOMAIN: The spacer domain and the TSP type-1 domains are important  
 CC for a tight interaction with the extracellular matrix.  
 CC -!- PMW: The precursor is cleaved by a furin endopeptidase (By  
 CC similarity).  
 CC -!- SIMILARITY: Belongs to peptidase family M12B.  
 CC -!- SIMILARITY: Contains 1 disintegrin-like domain.  
 CC -!- SIMILARITY: Contains 1 PLAC domain.  
 CC -!- SIMILARITY: Contains 4 TSP type-1 domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial









```

FT DOMAIN 572 700 CYS-RICH.
FT DOMAIN 701 838 SPACER.
FT DOMAIN 839 895 TSP TYPE-1 2.
FT DOMAIN 896 949 TSP TYPE-1 3.
FT SITE 174 174 CYS-STEIN SWITCH (POTENTIAL).
FT METAL 361 361 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 362 362 BY SIMILARITY.
FT METAL 365 365 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 371 371 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 591 591 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 679 679 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 950 AA; 103286 MW; 5DFBE-8285CCCC3B CRC64;

Query Match 13.4%; Score 111; DB 1; Length 950;
Best Local Similarity 33.0%; Pred. No. 0.0013;
Matches 31; Conservative 13; Mismatches 38; Indels 12; Gaps 6;

QY 44 WKPGPWGCTGDCGPGVQSRVAVCFHVDGWTSHLNSGSENNRPPKERSCFRVCMDHSDL 103
DB 843 WVAGSWGPGASCG-SGLQKRAVDCRSAGORT-VFACDAHRPVETQACGEPCP----- 895

QY 104 FQNEVSDWHHCVLPYARGVKPRTECVTAQH 137
DB 896 -TWELSAWSPC-SKSCRG-FQRLSKCV--GHG 924

RESULT 11
ID AT19 HUMAN STANDARD; PRT; 1207 AA.
AC Q8TE59;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE ADAMTS-19 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
with thrombospondin motifs 19) (ADAM-TS 19) (ADAM-TS19).
EN ADAMTS19.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ON NCBI_TaxID=9606;
RX MEDLINE=21856482; PubMed=11867212;
RA Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,
RA Lopez-Otin C.;
RT "Cloning, expression analysis, and structural characterization of
seven novel human ADAMTSs, a family of metalloproteinases with
disintegrin and thrombospondin-1 domains.";
RL Gene 283:49-62(2002).
CC -! COPACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -! SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
matrix (By similarity).
CC -! TISSUE SPECIFICITY: Expressed in fetal lung, but not in any adult
tissues examined. Expression was detected in an osteosarcoma CDNA
library.
CC -! PRT: The precursor is cleaved by a furin endopeptidase (By
similarity).
CC -! SIMILARITY: Belongs to peptidase family M12B.
CC -! SIMILARITY: Contains 1 disintegrin-like domain.
CC -! SIMILARITY: Contains 1 PLAC domain.
CC -! SIMILARITY: Contains 5 TSP type-1 domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ311904; CAC84565.1; -.

```

```

DR Genew; HGNC:17111; ADAMTS19.
DR MIM; 607513; -.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR001818; Pept_M10A_M12B.
DR InterPro; IPR006925; Pept_M_Zn_BS.
DR InterPro; IPR005990; Peptidase_M12B.
DR InterPro; IPR002870; Peptidase_M12B_N.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF01562; Pept_M12B_propep; 1.
DR Pfam; PF01421; Replicolysin; 1.
DR Pfam; PF00090; TSP1; 5.
DR SMART; SM00209; TSP1; 5.
DR PROSITE; PS0215; ADAM_MPEPRO; 1.
DR PROSITE; PS0546; CYSTEINE SWITCH; FALSE_NEG.
DR PROSITE; PS0427; DISINTEGRIN_1; FALSE_NEG.
DR PROSITE; PS0214; DISINTEGRIN_2; FALSE_NEG.
DR PROSITE; PS0900; PLAC; 1.
DR PROSITE; PS0092; TSP1; 5.
DR PROSITE; PS0142; ZINC_PROTEASE; 1.
DR K01; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix.
KW SIGNAL 1 21 POTENTIAL.
FT PROPEP 22 316 BY SIMILARITY.
FT CHAIN 317 1207 ADAMTS-19.
FT DOMAIN 317 545 METALLOPROTEASE.
FT DOMAIN 546 633 DISINTEGRIN-LIKE.
FT DOMAIN 634 686 TSP TYPE-1 1.
FT DOMAIN 687 790 CYS-RICH.
FT DOMAIN 791 914 SPACER.
FT DOMAIN 915 975 TSP TYPE-1 2.
FT DOMAIN 976 1037 TSP TYPE-1 3.
FT DOMAIN 1039 1083 TSP TYPE-1 4.
FT DOMAIN 1087 1144 TSP TYPE-1 5.
FT DOMAIN 1160 1199 PLAC.
FT DOMAIN 88 216 PRO-RICH.
FT SITE 294 294 CYSTEINE SWITCH (POTENTIAL).
FT METAL 482 482 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 483 483 BY SIMILARITY.
FT METAL 486 486 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 492 492 ZINC (CATALYTIC) (BY SIMILARITY).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 797 797 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 907 907 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 949 949 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1009 1009 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1207 AA; 134061 MW; AF36F6BF5886FDE2 CRC64;

Query Match 13.3%; Score 110.5; DB 1; Length 1207;
Best Local Similarity 27.3%; Pred. No. 0.0018;
Matches 30; Conservative 19; Mismatches 44; Indels 17; Gaps 7;

QY 31 AAHLGKKNQPIWKPGPWGCTGDCGPGVQSRVAVCFHVDGWTSHLNSGSENNRPPKE 90
DB 1028 AQRCQG-QDCMTVWEAGVWSFYSKCGK-GIRHRTVC-----TNPRKCVLSTRPRA 1079

QY 91 RSCFRVCDHSDLPQNEVSDWHHCVLPYARGVKPRTECVTAQH 137
DB 1080 ED---CEDYSKYVWRMGDSK-SITCGKG-MQSRVIOCMHKITGRHG 1123

RESULT 12
ATSI_RAT
ID ATSI_RAT STANDARD; PRT; 967 AA.
AC Q9WUQ1; Q9ERI1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE ADAMTS-1 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1).
GN ADAMTS1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

```

NCBI\_TaxID=10116;  
 [1]  
 SEQUENCE FROM N.A.  
 STRAIN-Sprague-Dawley; TISSUE-Brain;  
 Liu X., Tu Y., Yin T., Johnstone E.M., Stephenson D.T., Clemens J.A.,  
 Little S.P.;  
 "Induction of a disintegrin and metalloprotease with the  
 thrombospondin type I motif (ADAMTS).";  
 Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 [2]  
 SEQUENCE OF 18-967 FROM N.A.  
 STRAIN-Sprague-Dawley; TISSUE-Liver;  
 MEDLINE=20304099; PubMed=10847486;  
 Diamantis I., Luethi M., Hoelsi M., Reichen J.;  
 "Cloning of the rat ADAMTS-1 gene and its down regulation in  
 endothelial cells in cirrhotic rats.";  
 Liver 20:165-172(2000).  
 CC -!- FUNCTION: Cleaves aggrecan, a cartilage proteoglycan, and may be  
 involved in its turnover. Has angiogenic inhibitor activity (By  
 similarity). Active metalloprotease, which may be associated with  
 various inflammatory processes as well as development of cancer  
 cachexia. May play a critical role in follicular rupture (By  
 similarity).  
 CC -!- CATALYTIC ACTIVITY: Cleaves aggrecan at the 1683-Glu-|-Leu-1684  
 site, within the chondroitin sulfate attachment domain.  
 CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular  
 matrix (By similarity).  
 CC -!- INDUCTION: Down-regulated in endothelial cells derived from  
 cirrhotic liver.  
 CC -!- DOMAIN: The spacer domain and the TSP type-1 domains are important  
 for a tight interaction with the extracellular matrix.  
 CC -!- PTM: The precursor is cleaved by a furin endopeptidase (By  
 similarity).  
 CC -!- SIMILARITY: Belongs to peptidase family M12B.  
 CC -!- SIMILARITY: Contains 1 disintegrin-like domain.  
 CC -!- SIMILARITY: Contains 3 TSP type-1 domains.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 EMBL; AF149118; RAD34012.1; -;  
 EMBL; AF304446; AAG29823.1; -;  
 MEROPS; M12.222; -;  
 InterPro; IPR006586; ADAM\_cysteine.  
 InterPro; IPR001762; Disintegrin.  
 InterPro; IPR006025; Pept\_M\_Zn\_BS.  
 InterPro; IPR001590; Peptidase\_M12B.  
 InterPro; IPR002870; Peptidase\_M12B\_N.  
 InterPro; IPR000884; TSP1.  
 InterPro; IPR008085; TSP 1.  
 Pfam; PF01562; Pep\_M12B\_propep; 1.  
 Pfam; PF01421; Reprolysin; 1.  
 Pfam; PFC0090; tsp 1; 3.  
 PRINTS; PR01705; TSP1REPEAT.  
 SMART; SM00608; ACR; 1.  
 SMART; SM00209; TSP1; 3.  
 PROSITE; PS00215; ADAM\_MEPPO; 1.  
 PROSITE; PS00427; DISINTEGRIN\_1; FALSE\_NEG.  
 PROSITE; PS00214; DISINTEGRIN\_2; FALSE\_NEG.  
 PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 PROSITE; PS00092; TSP1; 3.  
 Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;  
 Repeat; Extracellular matrix; Heparin-binding.  
 SIGNAL  
 PROPEP 1 54 POTENTIAL.  
 CHAIN 55 252 BY SIMILARITY.  
 CHAIN 253 967 ADAMTS-1.  
 DOMAIN 253 475 METALLOPROTEASE.

DOMAIN 476 558 DISINTEGRIN-LIKE.  
 TSP TYPE-1 1.  
 CYS-RICH.  
 SPACER.  
 TSP TYPE-1 2.  
 TSP TYPE-1 3.  
 POLY-ARG.  
 CYSTEINE SWITCH (POTENTIAL).  
 ZINC (CATALYTIC) (BY SIMILARITY).  
 BY SIMILARITY.  
 ZINC (CATALYTIC) (BY SIMILARITY).  
 ZINC (CATALYTIC) (BY SIMILARITY).  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 I -> V (IN REF. 2).  
 KERSQ -> RSGSL (IN REF. 2).  
 V -> A (IN REF. 2).  
 R -> P (IN REF. 2).  
 L -> TR (IN REF. 2).  
 R -> G (IN REF. 2).  
 TMLV -> NLK (IN REF. 2).  
 S -> F (IN REF. 2).  
 L -> V (IN REF. 2).  
 I -> T (IN REF. 2).  
 SQ SEQUENCE 967 AA; 105705 MW; F93C864F6DCB4CF CRC64;  
 Query Match 13.3%; Score 110; DB 1; Length 967;  
 Best Local Similarity 27.9%; Pred. No. 0.0016;  
 Matches 36; Conservative 17; Mismatches 48; Indels 28; Gaps 8;  
 QY 26 LLSHA-----AHLEKKNQF-----TWKPGWRCCTGDCGPGVQSRAYWCFHV 71  
 DB 826 LVGHALRPKIKYTFMKKKTFEPAFTFSEWIEWGECKTCG-SGQRRVVECRDI 884  
 QY 72 DGTWTHSLNCGESNRPPKERSFRVCDMSDF--PQWVSDWHHCVLVYARGEVKPRTA 129  
 DB 885 NGHPA--SECAKEVKPASTRPC-----ADLPGRWQVGDWSPCKT-CGKG-YKCRTL 933  
 QY 130 ECVTAQHL 138  
 DB 934 KCLSHDGV 942  
 RESULT 13  
 ATSP\_HUMAN STANDARD; PRT; 1211 AA.  
 ID ATSP\_HUMAN  
 AC O95450;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE ADAMTS-2 precursor (EC 3.4.24.14) (A disintegrin and metalloprotease  
 with thrombospondin motifs 2) (ADAM-TS 2) (ADAM-TS2) (Procollagen I/II  
 amino-propeptide processing enzyme) (Procollagen I N-proteinase) (PC  
 I-NP) (Procollagen N-endopeptidase) (pNPI).  
 GN ADAMTS2 OR PCINP OR PCPNI.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OC NCBI\_TaxID=9606;  
 OX [1]  
 RN SEQUENCE FROM N.A. (ISOFORMS LNP1 AND SPNP1), AND, DISEASE.  
 RC TISSUE=Skin;  
 RC MEDLINE=99347935; PubMed=10417273;  
 RA Colige A., Sieron A.L., Li S.-W., Schwarze U., Petty E.,  
 RA Wartecki W., Wilcox W., Krakow D., Cohn D.H., Reardon W.,  
 RA Byers P.H., Lapierre C.M., Prockop D.J., Nussgens B.V.;  
 RT "Human Ehlers-Danlos syndrome type VII C and bovine dermatosparaxis  
 are caused by mutations in the procollagen I N-proteinase gene.";  
 RL Am. J. Hum. Genet. 65:308-317(1999).  
 CC -!- FUNCTION: Cleaves the propeptides of type I and II collagen prior

to fibril assembly. Does not act on type III collagen. May also play a role in development that is independent of its role in collagen biosynthesis.

-!- CATALYTIC ACTIVITY: Cleaves the N-propeptide of collagen chain at Ala-I(I)-at Pro-I-Gln and of alpha-1(I) and alpha-2(I) chains at Ala-I-Gln.

-!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).

-!- SUBUNIT: May belong to a multimeric complex. Binds specifically to collagen type XIV (By similarity).

-!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (By similarity).

-!- ALTERNATIVE PRODUCTS:

Event=Alternative splicing; Named isoforms=2;

Name=LGNPI;

isoId=O95450-1; Sequence=Displayed;

Name=SPNPI;

isoId=O95450-2; Sequence=VSP\_005497, VSP\_005498;

Note=Has no significant N-procollagen peptidase activity;

-!- TISSUE SPECIFICITY: Expressed at high level in skin, bone, tendon and aorta and at low levels in thymus and brain.

-!- DOMAIN: The spacer domain and the TSP type-1 domains are important for a tight interaction with the extracellular matrix.

-!- PTM: The precursor is cleaved by a furin endopeptidase (By similarity).

-!- DISEASE: Defects in ADAMTS2 are the cause of Ehlers-Danlos syndrome type VIIC (EDS VIIC) [MIM:225410]. EDS VIIC is a recessively inherited connective-tissue disorder characterized clinically by severe skin fragility and joint hypermobility and biochemically by the presence in skin of procollagen incompletely processed at the N-terminus.

-!- SIMILARITY: Belongs to peptidase family M12B.

-!- SIMILARITY: Contains 1 disintegrin-like domain.

-!- SIMILARITY: Contains 4 TSP domain.

-!- CAUTION: Has sometimes been referred to as ADAMTS3.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: AJ003125; CAA05880.1; -

MEKOPS; M12.301; -

Genew; HGNC:218; ADAMTS2.

MIM; 604539; -

MIM; 225410; -

GO; GO:0008133; F:collagenase activity; TAS.

GO; GO:0008237; F:metallopeptidase activity; TAS.

InterPro; IPR001762; Disintegrin.

InterPro; IPR006025; Pept\_Mn\_BS.

InterPro; IPR001590; Peptidase\_M12B.

InterPro; IPR002870; Peptidase\_M12B\_N.

InterPro; IPR000884; TSP1.

InterPro; IPR008095; TSP1.

Pfam; PF01562; Pep\_M12B\_Propag; 1.

Pfam; PF01421; Repolysin; 1.

Pfam; PF00090; tsp\_1; 4.

PRINTS; PR01705; TSP1REPEAT.

SMART; SM00209; TSP1; 4.

PROSITE; PS02015; ADAM\_MEPPO; 1.

PROSITE; PS00427; DISINTEGRIN\_1; FALSE NEG.

PROSITE; PS02014; DISINTEGRIN\_2; FALSE\_NEG.

PROSITE; PS00900; PLAC; 1.

PROSITE; PS00092; TSP1; 4.

PROSITE; PS00142; ZINC\_PROTEASE; FALSE NEG.

PROSITE; PS00142; ZINC\_PROTEASE; FALSE NEG.

KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;

Repeat; Collagen degradation; Extracellular matrix;

Alternative splicing; Ehlers-Danlos syndrome.

SIGNAL 1 29 POTENTIAL.

PROPEP 30 253 BY SIMILARITY.

FT CHAIN 254 1211 ADAMTS-2.

FT DOMAIN 254 479 METALLOPROTEASE.

FT DOMAIN 480 560 DISINTEGRIN-LIKE.

FT DOMAIN 561 616 TSP TYPE-1 1.

FT DOMAIN 618 722 CYS-RICH.

FT SITE 691 693 CELL ATTACHMENT SITE (POTENTIAL).

FT DOMAIN 723 851 SPACER.

FT DOMAIN 854 912 TSP TYPE-1 2.

FT DOMAIN 914 971 TSP TYPE-1 3.

FT DOMAIN 975 1029 TSP TYPE-1 4.

FT DOMAIN 1059 1097 PLAC.

FT DOMAIN 140 43 POLY-ALA.

FT METAL 185 188 POLY-GLU.

FT METAL 408 408 ZINC (CATALYTIC) (BY SIMILARITY).

FT ACT SITE 409 409 ZINC (CATALYTIC) (BY SIMILARITY).

FT METAL 412 412 ZINC (CATALYTIC) (BY SIMILARITY).

FT METAL 418 418 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 251 251 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 949 949 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 993 993 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1031 1031 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1098 1098 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1145 1145 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1150 1150 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT VARSPLIC 544 566 HCPKGRHCWLPDLKKDGSNGA -> FRGAVAHACYPST

FT LGGQGRWIA (in isoform SpNP1).

FT Missing (in isoform SpNP1).

FT /FtId=VSP\_005497.

FT /FtId=VSP\_005498.

SQ SEQUENCE 1211 AA; 134722 MW; BECEEFE25C23CAD2D CRC64;

Query Match 13.2%; Score 109.5; DB 1; Length 1211;

Best Local Similarity 28.0%; Pred. No. 0.0023;

Matches 30; Conservative 12; Mismatches 38; Indels 27; Gaps 6;

QY 40 NQFTWKCPGRCCTGDCGPGVQSRVAVMCFHV--DGWTS--HLSNCGESNPPKRCSCFR 95

Db 914 SQPVVVTGWEPCSTCGRTGMQVRSVRCIPLHNTTRSHVAKHCNDA-RPESRRACSR 972

QY 96 -VCDHSDLFQWVSDVHWHCVLPVYARGEVKPRTAEVCVTAHQGLQHR 141

Db 973 ELCPG-----RWAGPWSQC-----SVTCGNGTQER 998

RESULT 14

AT10\_HUMAN

ID AT10\_HUMAN STANDARD; PRT; 1077 AA.

AC Q9H324;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE ADAMTS-10 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 10) (ADAM-TS 10) (ADAM-TS10) (Fragment).

GN ADAMTS10.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]\_TaxID=9606;

RP SEQUENCE FROM N.A.

RA Apte S.S.;

RT "ADAM-TS10; a novel member of the ADAM-TS family containing multiple thrombospondin type I repeats.";

RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.

CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).

CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (By similarity).

CC -!- DOMAIN: The spacer domain and the TSP type-1 domains are important for a tight interaction with the extracellular matrix (By similarity).

CC -!- SIMILARITY: Belongs to peptidase family M12B.

CC -!- SIMILARITY: Contains 1 disintegrin-like domain.

CC -1- SIMILARITY: Contains 1 PLAC domain.  
 CC -1- SIMILARITY: Contains 5 TSP type-1 domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF163762; AAC35563.1; -  
 CC DR MEROPS; M12.235; -  
 CC DR Genew; HGNC:13201; ADAMTS10.  
 CC DR InterPro; IPR001762; Disintegrin.  
 CC DR InterPro; IPR006025; Pept\_M\_2n\_BS.  
 CC DR InterPro; IPR001590; Peptidase\_M12B.  
 CC DR InterPro; IPR002870; Peptidase\_M12B\_N.  
 CC DR InterPro; IPR000884; TSP1.  
 CC DR InterPro; IPR008085; TSP\_1.  
 CC DR Pfam; PF01562; Pep\_M12B\_propep; 1.  
 CC DR Pfam; PF01421; Reptolysin\_1.  
 CC DR Pfam; PF00090; TSP\_1\_5.  
 CC DR PRINTS; PR01705; TSP1REPEAT.  
 CC DR SMART; SM00209; TSP1; 5.  
 CC DR PROSITE; PS00215; ADAM\_MEPRO; 1.  
 CC DR PROSITE; PS00427; DISINTEGRIN\_1; FALSE\_NEG.  
 CC DR PROSITE; PS02114; DISINTEGRIN\_2; FALSE\_NEG.  
 CC DR PROSITE; PS05900; PLAC; 1.  
 CC DR PROSITE; PS00092; TSP1; 5.  
 CC DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 CC DR Hydroxylase; Metalloprotease; Zinc; Glycoprotein; Zymogen;  
 CC Repeat; Extracellular matrix.  
 CC KW NON TER 1  
 CC FT PROPEP <1 207 BY SIMILARITY.  
 CC FT CHAIN 208 1077 ADAMTS-10.  
 CC FT DOMAIN 208 433 METALLOPROTEASE.  
 CC FT DOMAIN 434 520 DISINTEGRIN-LIKE.  
 CC FT DOMAIN 578 679 CIS-RICH.  
 CC FT DOMAIN 680 802 SPACER.  
 CC FT DOMAIN 521 576 TSP TYPE-1 1.  
 CC FT DOMAIN 799 857 TSP TYPE-1 2.  
 CC FT DOMAIN 858 919 TSP TYPE-1 3.  
 CC FT DOMAIN 921 975 TSP TYPE-1 4.  
 CC FT DOMAIN 977 1032 TSP TYPE-1 5.  
 CC FT DOMAIN 1039 1077 PLAC.  
 CC FT METAL 366 366 ZINC (CATALYTIC) (BY SIMILARITY).  
 CC FT ACT SITE 367 367 ZINC (CATALYTIC) (BY SIMILARITY).  
 CC FT METAL 370 370 ZINC (CATALYTIC) (BY SIMILARITY).  
 CC FT METAL 376 376 ZINC (CATALYTIC) (BY SIMILARITY).  
 CC FT CARBOHYD 64 64 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC FT CARBOHYD 714 714 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC FT CARBOHYD 769 769 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC FT CARBOHYD 866 866 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC SQ SEQUENCE 1077 AA; 118072 MW; 3914DEL8DCBFF587 CRC64;  
 CC  
 CC Query Match 13.1%; Score 108.5; DB 1; Length 1077;  
 CC Best Local Similarity 41.8%; Pred. No. 0.0026;  
 CC Matches 23; Conservative 4; Mismatches 25; Indels 3; Gaps 2;  
 CC  
 CC QY 44 WKPGWGRTGCGGQGVQSRVAVGCFHYDGTWTSLSNGCESNRPPKSCRFVCD 98  
 CC DB 981 WAGENGECACQGVGQGRQ-RSVRCTSHGTQASH--ECTEALRPPTTQCCAKCD 1032  
 CC  
 CC RESULT 15  
 CC AT2S\_BOVIN STANDARD; PRT; 1205 AA.  
 CC AC P79331;  
 CC DT 16-OCT-2001 (Rel. 40, Created)  
 CC DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE ADAMTS-2 precursor (EC 3.4.24.14) (A disintegrin and metalloproteinase  
 DE with thrombospondin motifs 2) (ADAM-TS 2) (ADAM-TS2) (Procollagen I/II  
 DE amino-propeptide processing enzyme) (Procollagen I N-proteinase) (PC  
 GN I-NP) (Procollagen N-endopeptidase) (PNPI).  
 GN ADAMTS2 OR NPI.  
 OC Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OC NCBI\_TaxID=9913;  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC TISSUE=Skin;  
 CC MEDLINE=97225960; PubMed=9122202;  
 CC Collige A., Li S.W., Sieron A.L., Nusgens B.V., Prockop D.J.,  
 CC Lapiere C.M.;  
 CC "cDNA cloning and expression of bovine procollagen I N-proteinase: a  
 CC new member of the superfamily of zinc-metalloproteinases with binding  
 CC sites for cells and other matrix components.";  
 CC Proc. Natl. Acad. Sci. U.S.A. 94:2374-2379 (1997).  
 CC [2]  
 CC PARTIAL SEQUENCE.  
 CC MEDLINE=95348096; PubMed=7622483;  
 CC Collige A., Beschin A., Samyn B., Goebels Y., Van Beeumen J.,  
 CC Nusgens B.V., Lapiere C.M.;  
 CC "Characterization and partial amino acid sequencing of a 107-kDa  
 CC procollagen I N-proteinase purified by affinity chromatography on  
 CC immobilized type XIV collagen.";  
 CC J. Biol. Chem. 270:16724-16730 (1995).  
 CC -1- FUNCTION: Cleaves the propeptides of type I and II collagen prior  
 CC to fibril assembly. Does not act on type III collagen. May also  
 CC play a role in development that is independent of its role in  
 CC collagen biosynthesis.  
 CC -1- CATALYTIC ACTIVITY: Cleaves the N-propeptide of collagen chain  
 CC alpha-1(I) at Pro-[Gln and of alpha-1(II) and alpha-2(I) chains  
 CC at Ala-[Gln.  
 CC -1- COFACTOR: Binds 1 zinc ion per subunit (By similarity).  
 CC -1- SUBUNIT: May belong to a multimeric complex. Binds specifically to  
 CC collagen type XIV.  
 CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular  
 CC matrix (By similarity).  
 CC -1- TISSUE SPECIFICITY: Enzymatic activity is detected at high level  
 CC in all type I collagen-rich tissues such as skin, bones, tendons  
 CC and aorta and at low level in brain and thymus. The mRNA levels  
 CC were disproportionately high in heart, liver, retina and muscle.  
 CC -1- DOMAIN: The spacer domain and the TSP type-1 domains are important  
 CC for a tight interaction with the extracellular matrix.  
 CC -1- PTM: The N-terminus is blocked.  
 CC -1- PTM: The precursor is cleaved by a furin endopeptidase (By  
 CC similarity).  
 CC -1- DISEASE: Defects in ADAMTS2 are the cause of dermatosparaxis, a  
 CC recessively inherited disorder characterized by severe skin  
 CC fragility and biochemically by the presence in skin of procollagen  
 CC incompletely processed at the N-terminus.  
 CC -1- SIMILARITY: Belongs to peptidase family M12B.  
 CC -1- SIMILARITY: Contains 1 disintegrin-like domain.  
 CC -1- SIMILARITY: Contains 1 PLAC domain.  
 CC -1- SIMILARITY: Contains 4 TSP type-1 domains.  
 CC -1- CAUTION: Has sometimes been referred to as ADAMTS3.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X96389; CAA65253.1; -  
 CC DR PIR; T18517; T18517.  
 CC DR MEROPS; M12.301; -  
 CC DR InterPro; IPR001762; Disintegrin.

```

OR InterPro; IPR006025; Pept_M_Zn_BS.
OR InterPro; IPR001590; Peptidase_M12a.
OR InterPro; IPR002870; Peptidase_M12b_N.
OR InterPro; IPR000884; TSP1.
OR Pfam; PF01362; Pep_M12B_propep; 1.
OR Pfam; PF01421; Reprolysin; 1.
OR Pfam; PF00090; tsp; 1; 4.
OR SMART; SM00209; TSP1; 4.
OR PROSITE; PS00215; ADAM_MEPPO; 1.
OR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
OR PROSITE; PS00214; DISINTEGRIN_2; FALSE_NEG.
OR PROSITE; PS00300; PLAC; 1.
OR PROSITE; PS00092; TSP1; 4.
OR PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
OR Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
OR Repeat; Collagen degradation; Extracellular matrix.
OR SIGNAL 1 28 POTENTIAL.
OR PROPEP 29 253 BY SIMILARITY.
OR CHAIN 254 1205 ADAMS-2.
OR DOMAIN 254 473 METALLOPROTEASE.
OR DOMAIN 474 554 DISINTEGRIN-LIKE.
OR DOMAIN 555 610 TSP TYPE-1 1.
OR DOMAIN 612 716 CYS-RICH.
OR DOMAIN 717 845 SPACER.
OR DOMAIN 848 906 TSP TYPE-1 2.
OR DOMAIN 908 968 TSP TYPE-1 3.
OR DOMAIN 969 1023 TSP TYPE-1 4.
OR DOMAIN 1053 1091 PLAC.
OR DOMAIN 31 35 POLY-ALA.
OR DOMAIN 177 180 POLY-GLU.
OR METAL 402 402 ZINC (CATALYTIC) (POTENTIAL).
OR ACT_SITE 403 403 BY SIMILARITY.
OR METAL 406 406 ZINC (CATALYTIC) (BY SIMILARITY).
OR METAL 412 412 ZINC (CATALYTIC) (BY SIMILARITY).
OR SITE 685 687 CELL ATTACHMENT SITE (POTENTIAL).
OR CARBOHYD 104 104 N-LINKED (GLCNAC. .) (POTENTIAL).
OR CARBOHYD 245 245 N-LINKED (GLCNAC. .) (POTENTIAL).
OR CARBOHYD 942 942 N-LINKED (GLCNAC. .) (POTENTIAL).
OR CARBOHYD 943 943 N-LINKED (GLCNAC. .) (POTENTIAL).
OR CARBOHYD 987 987 N-LINKED (GLCNAC. .) (POTENTIAL).
OR CARBOHYD 1025 1025 N-LINKED (GLCNAC. .) (POTENTIAL).
OR CARBOHYD 1092 1092 N-LINKED (GLCNAC. .) (POTENTIAL).
OR CARBOHYD 1139 1139 N-LINKED (GLCNAC. .) (POTENTIAL).
OR CARBOHYD 1144 1144 N-LINKED (GLCNAC. .) (POTENTIAL).
OR SEQUENCE 1205 AA; 133887 MW; 735B232A45320371 CRC64;

Query Match 13.0%; Score 107.5; DB 1; Length 1205;
Best Local Similarity 28.0%; Pred. No. 0.0037;
Matches 30; Conservative 13; Mismatches 37; Indels 27; Gaps 6;

2y 40 NQFIWKPGKGRCTGCGPGGVQSVAVWC---PHVDGWTSLHLSNCGESNRPPKERSQPR 95
Db 908 SQPVWVTGEWEPSCRGRTGMQVRSCVQPLHNTTTSVHTKCNDA-RPEGRACNR 966

2y 96 -VCDHSDLFQWEVSDWHHCVLVPYARGEYKPTACVTAQHGQLQR 141
Db 967 ELCPG-----RWAGWSQC-----SVTCNGTQER 992
```

Search completed: February 24, 2004, 01:03:08  
Job time : 4.97635 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM protein - protein search, using sw model

Run on: February 24, 2004, 00:30:51 ; Search time 10.692 Seconds  
(without alignments)  
4190.395 Million cell updates/sec

Title: US-10-022-710-6\_COPY\_1\_142

Perfect score: 830

Sequence: 1 MFPKSLNLTVCWRSWRKL.....EVKPTAEVCVTAQHGLOHRM 142

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL: 25:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phage:\*

10: sp\_plant:\*

11: sp\_protist:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_rviro:\*

16: sp\_bacteriopl:\*

17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	667	80.4	1107	11 Q8BHP3	Q8bhp3 mus musculus
2	414	49.9	1536	4 Q9C014	Q9c014 homo sapien
3	127	15.3	622	4 Q9H8X0	Q9h8x0 homo sapien
4	126	15.2	1280	11 Q9EPX2	Q9epx2 mus musculus
5	125	15.1	701	11 Q8CDB8	Q8cdb8 mus musculus
6	124	14.9	711	5 Q8I411	Q8i411 caenorhabdi
7	121.5	14.6	957	11 Q7TSK7	Q7tsk7 mus musculus
8	120.5	14.5	1235	4 Q95428	Q95428 homo sapien
9	119.5	14.4	951	4 Q60345	Q60345 homo sapien
10	119.5	14.4	951	4 Q86TH1	Q86th1 homo sapien
11	118	14.2	724	5 Q15916	Q15916 dictyosteli
12	117.5	14.2	898	4 Q9UFZ4	Q9ufz4 homo sapien
13	116.5	14.0	340	11 Q91Z56	Q91z56 mus musculus
14	116	14.0	237	4 Q9H8S6	Q9h8s6 homo sapien
15	116	14.0	424	4 Q8N643	Q8n643 homo sapien
16	114.5	13.8	640	11 Q8BSM2	Q8bsm2 mus musculus

17	114.5	13.8	1043	5 P30884	P30884 caenorhabdi
18	112	13.5	235	11 Q91YP4	Q91yp4 mus musculus
19	112	13.5	1036	11 Q80T21	Q80t21 mus musculus
20	111.5	13.4	1502	4 Q9UP26	Q9up26 homo sapien
21	111.5	13.3	1092	21 Q8BXA1	Q8bka1 mus musculus
22	110.5	13.2	2165	5 Q19751	Q19751 caenorhabdi
23	109	13.1	1023	4 Q9ULI7	Q9uli7 homo sapien
24	107	12.9	759	6 Q8H2M8	Q8hzm8 equus cabal
25	106.5	12.8	446	4 Q43384	Q43384 homo sapien
26	106.5	12.8	740	4 Q8WUV0	Q8wuv0 homo sapien
27	105.5	12.7	310	11 Q8C9R1	Q8c9r1 mus musculus
28	105	12.7	3198	5 Q9U8G8	Q9u8g8 manduca sex
29	103.5	12.5	432	4 Q9NPM2	Q9npm2 homo sapien
30	103.5	12.5	2174	5 Q9GQK0	Q9gqk0 drosophila
31	103.5	12.5	2772	5 Q9AV4	Q9av4 drosophila
32	103.5	12.5	2776	5 Q869A0	Q869a0 drosophila
33	103.5	12.5	2898	5 Q868Z9	Q868z9 drosophila
34	103	12.4	417	4 Q9BXV3	Q9bxv3 homo sapien
35	103	12.4	525	4 Q96RW4	Q96rw4 homo sapien
36	103	12.4	1487	5 Q8MPV5	Q8mpv5 caenorhabdi
37	103	12.4	1558	5 Q8I710	Q8i710 caenorhabdi
38	103	12.4	2167	5 Q76840	Q76840 caenorhabdi
39	102.5	12.3	364	4 Q96AY5	Q96ay5 homo sapien
40	101	12.2	538	4 Q9H8E4	Q9h8e4 homo sapien
41	100	12.0	1009	11 Q8BKY1	Q8bky1 mus musculus
42	100	12.0	1600	11 Q811B3	Q811b3 mus musculus
43	99.5	12.0	377	11 Q8BLI0	Q8bli0 mus musculus
44	99.5	12.0	761	6 Q95LQ2	Q95lq2 macaca fasc
45	98.5	11.9	192	11 Q9CX59	Q9cx59 mus musculus

ALIGNMENTS

RESULT 1

Q8BHP3 ID Q8BHP3 PRELIMINARY; PRT; 1107 AA.  
AC Q8BHP3;  
DT 01-MAR-2003 (TRENBLrel. 23, Created)  
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE Weakly similar to hypothetical protein DJ110N13.1 in chromosome 7.  
DE 7.  
GN D130067103RIK.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Dorsal root ganglion;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium,  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs."  
RL Nature 420:563-573 (2002).  
DR EMBL; AK051714; BAC34732.1; -  
DR MGD; MGI:2443925; D130067103RIK.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0004691; F:oxidoreductase activity; IEA.  
DR GO; GO:0006118; P:electron transport; IEA.  
DR InterPro; IPR005797; Cytb\_b6\_N.  
DR InterPro; IPR000884; TSP1.  
DR Pfam; PF00090; TSP1; 7.  
DR SMART; SM00209; TSP1; 9.  
DR PROSITE; PS00192; CYTOCHROME\_B\_HEME; 1.  
DR PROSITE; PS00092; TSP1; 10.  
KW Hypothetical protein.  
SQ SEQUENCE 1107 AA; 123640 MW; B34C708E6AA65734 CRC64;  
Query Match 80.4%; Score 667; DB 11; Length 1107;  
Best Local Similarity 80.1%; Pred. No. 2.7e-67;



Matches 113; Conservative 13; Mismatches 15; Indels 0; Gaps 0;

QY 1 MFPSNLTVCWVRSWEKFLLLSLLSHAHLEGGKDNQFIWPGPWGRCTGDCGPGG 60  
 DB 1 MFLKSDLAIVHWRSMEKLFVLJLSLSQAHLGRKDNQFLWKTGPWGRCAAGDCGPGG 60

QY 61 VQSRVAVCFHVDGWTSHLSNGESNRPPKERSCFRVCDDHSDLFQWEVSDWHHCVLVPPYA 120  
 DB 61 AQSRVAVCFHIEGWTSPMSNDESSQPKERSCFRVCDDHSDLFQWEVSDWHHCVLVPPYA 120

QY 121 RGEVPRTAECVTAQHGLQHR 141  
 DB 121 QGEPRPRAVECVTAQHGLQHR 141

## RESULT 2

QSC014  
 ID Q9C014 PRELIMINARY; PRT; 1536 AA.

AC Q9C014;  
 DT 01-JUN-2001 (TREMELrel. 17, Created)  
 DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)  
 DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)  
 DE Hypothetical protein KIAA1679 (Fragment).  
 GN KIAA1679.

OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE=21082932; PubMed=11214970;  
 RA Nagase T., Kikuno R., Hattori A., Kondo Y., Okumura K., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. XIX.  
 RT The complete sequences of 100 new cDNA clones from brain which code  
 RT for large proteins in vitro."  
 RL DNA Res. 7:347-355(2000).  
 DR EMBL; AB051466; BAB21770.1; --  
 DR InterPro; IPR000884; TSP1.  
 DR Pfam; PF00090; tsp.1; 8.  
 DR SMART; SM00209; TSP1; 16.  
 DR PROSITE; PS00992; TSP1; 12.  
 KW Hypothetical protein.  
 FT NON TER 1  
 SQ SEQUENCE 1536 AA; 171242 MW; B5121C8160ECEDCE CRC64;

Query Match 49.9%; Score 414; DB 4; Length 1536;  
 Best Local Similarity 100.0%; Pred. No. 38-38;  
 Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 GWTSHLSNGESNRPPKERSCFRVCDDHSDLFQWEVSDWHHCVLVPPYARGEVKPRTAECV 132  
 DB 1 GWTSHLSNGESNRPPKERSCFRVCDDHSDLFQWEVSDWHHCVLVPPYARGEVKPRTAECV 60

QY 133 TAQHGLQHRM 142  
 DB 61 TAQHGLQHRM 70

## RESULT 3

Q9H8X0  
 ID Q9H8X0 PRELIMINARY; PRT; 622 AA.

AC Q9H8X0;  
 DT 01-MAR-2001 (TREMELrel. 16, Created)  
 DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)  
 DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)  
 DE Hypothetical protein FLJ13166.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RA Isogai T., Ota T., Hayaashi K., Sugiyama T., Otsuki T., Suzuki Y.,

RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,  
 RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,  
 RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,  
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,  
 RA Ninomiya K., Iwayanagi T.;  
 RT "NECD human cDNA sequencing project."  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK023228; BAB14479.1; --  
 DR InterPro; IPR000884; TSP1.  
 DR Pfam; PF00090; tsp.1; 6.  
 DR SMART; SM00209; TSP1; 8.  
 DR PROSITE; PS00992; TSP1; 7.  
 KW Hypothetical protein.  
 SQ SEQUENCE 622 AA; 70229 MW; DC415F663E5B8838 CRC64;

Query Match 15.3%; Score 127; DB 4; Length 622;  
 Best Local Similarity 29.7%; Pred. No. 7.1e-06;  
 Matches 30; Conservative 15; Mismatches 38; Indels 18; Gaps 5;

QY 39 DNQFI-----WKEGPMWRCTGCGPGVQSVRAVWCFHVDGWTSHLSNGESNRPPKE 90  
 DB 413 ENQVHPVSRGNCWRTGPGWGCSSCS--GGLQRAVVCQDENGQSA--SYCDAASKPPEL 469

QY 91 RSCFRVCDHSDLFQWEVSDWHHCVLVPPYARGEVKPRTAEC 131  
 DB 470 QQC-----GPGCPQWNYGNWGCESQT--CGGKISRLVIC 503

RESULT 4

Q9EPX2  
 ID Q9EPX2 PRELIMINARY; PRT; 1280 AA.

AC Q9EPX2;  
 DT 01-MAR-2001 (TREMELrel. 16, Created)  
 DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)  
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
 DE Papilin.  
 GN PAPLN.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=CD-1.  
 RX MEDLINE=20530499; PubMed=11076767;  
 RA Kramarova I.A., Kawaguchi N., Nelson R.E., Fessler L.I., Chen Y.,  
 RA Kramarov A.A., Kusche-Gullberg M., Kramer J.M., Ackley B.D.,  
 RA Sieron A.L., Prockop D.J., Fessler J.H.;  
 RT "Papilin in development; a pericellular protein with a homology to the  
 RT ADAMTS metalloproteinases."  
 RL Development 127:5475-5485(2000).  
 DR EMBL; AF314171; AAG41980.1; --  
 DR HSSP; P12111; 2KNT.  
 DR MGD; MGI:2386139; Papln.  
 DR GO; GO:0003824; F: catalytic activity; IEA.  
 DR GO; GO:0004867; F: serine protease inhibitor activity; IEA.  
 DR InterPro; IPR000183; Decarboxylase2.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003598; IG c2.  
 DR InterPro; IPR002223; Kunitz\_BPTI.  
 DR InterPro; IPR000884; TSP1.  
 DR Pfam; PF00047; IG; 3.  
 DR Pfam; PF00014; Kunitz\_BPTI; 1.  
 DR Pfam; PF00090; tsp.1; 5.  
 DR PRINTS; PR00759; BASICPTASE.  
 DR ProDom; PD000222; Kunitz\_BPTI; 1.  
 DR SMART; SM00408; IGC2; 3.  
 DR SMART; SM00131; KU; 1.  
 DR SMART; SM00209; TSP1; 5.  
 DR PROSITE; PS00280; BPTI\_KUNITZ\_1; 1.  
 DR PROSITE; PS0279; BPTI\_KUNITZ\_2; 1.  
 DR PROSITE; PS0835; IG\_LIKE; 3.

```
DR PROSITE; PS00879; ODR_DC_2_2; 1.
DR PROSITE; PS50092; TSP1; 5.
KW Immunoglobulin domain; Protease inhibitor; Serine protease inhibitor.
SQ SEQUENCE 1280 AA; 138824 MW; A2827705E561AF30 CRC64;

Query Match 15.2%; Score 126; DB 11; Length 1280;
Best Local Similarity 32.6%; Pred. No. 2e-05;
Matches 31; Conservative 15; Mismatches 39; Indels 10; Gaps 6;

QY 42 FIKPGGWRCTGDCGGVQSVRAVWC-FHVDGWTSHLSNCGSNRPPKERSC-FRVCDW 99
DB 307 FSHSGSWGDCSAECG-GGHSRLVFTINEXYDPHM--CQHPRTTHERSCNTQPC-- 361
QY 100 HSDLFQWESVDMHCVLPVARGVKPRTAECVTA 134
DB 362 -PKTKRWKVGFWTPCSV--SCGGGVQSVRSVYCISS 393

RESULT 5
Q8CDB8 PRELIMINARY; PRT; 701 AA.
ID Q8CDB8
AC Q8CDB8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein.
GN 8430403M15RIK OR AW743315.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=22354683; PubMed=12466851;
RA THE FANTOM Consortium,
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL; AK030699; BAC27087.1; -.
DR MGD; MGI:2141719; 8430403M15RIK.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00090; tsp.1; 8.
DR SMART; SM00209; TSP1; 8.
DR PROSITE; PS50092; TSP1; 8.
KW Hypothetical protein.
SQ SEQUENCE 701 AA; 77924 MW; C854E56D0704805F CRC64;

Query Match 15.1%; Score 125; DB 11; Length 701;
Best Local Similarity 31.5%; Pred. No. 1.4e-05;
Matches 29; Conservative 14; Mismatches 37; Indels 12; Gaps 5;

QY 44 WKPGWRCCTGDCGGVQSVRAVWC-FHVDGWTSHLSNCGSNRPPKERSC-FRVCDW 102
DB 125 WRTGFWGACSTCA-GGQRRVVVCDENGYTA--NDCVERIKPDEQACSGPCF---- 177
QY 103 LFQWESVDMHCVLPVARGVKPRTAECVTA 134
DB 178 --QWYSGWGECKL--CGGMRTRLVVCCRA 205

RESULT 6
Q8I4I1 PRELIMINARY; PRT; 711 AA.
ID Q8I4I1
AC Q8I4I1;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein F53B6.2b.
GN F53B6.2 OR F53B6.2b.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
```

```
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA White S.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018 (1998).
DR EMBL; Z81086; CAD56586.1; -.
DR WormPep; F53B6.2b; C332429.
DR InterPro; IPR003599; IG.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_C2.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00047; IG; 1.
DR Pfam; PF00090; tsp.1; 2.
DR SMART; SM00409; IG; 1.
DR SMART; SM00408; IGC2; 1.
DR SMART; SM00209; TSP1; 6.
DR PROSITE; PS50835; IG_Like; 1.
DR PROSITE; PS50092; TSP1; 5.
KW Hypothetical protein.
SQ SEQUENCE 711 AA; 79945 MW; 9695028B9636AE94 CRC64;

Query Match 14.9%; Score 124; DB 5; Length 711;
Best Local Similarity 29.5%; Pred. No. 1.8e-05;
Matches 38; Conservative 11; Mismatches 48; Indels 32; Gaps 6;

QY 17 MKXJFLLLSLLSHAHLE-----GKXDNQFIWK-----PGWRCCTGDCGG 59
DB 1 MLPLLLLSAPLGVSAFLDMNDRAYLDGNTFGSDMNDWQAPRLVAGWETCSSTCCT- 59
QY 60 GVQSVRAVWC-FHVDGWTSHLSNCGSNRPPKERSC-FRVCDW 105
DB 60 GWSRTVECVAVNPISAPIKLPWSECQDQPKLFSCCEVRSCLQEDSKLSDEAPYQ 119
QY 106 WEVSDWHHC 114
DB 120 WRYGDWTCQ 128

RESULT 7
Q7TSK7 PRELIMINARY; PRT; 957 AA.
ID Q7TSK7
AC Q7TSK7;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE TSP1-repeats containing protein.
GN TCP-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Spleen;
RA Kishi Y., Toji S., Tanaka M., Miyajima A., Yahara I.;
RT "Cloning of TSP1-repeats-containing protein expressed in splenic CD11c-
RT cells.";
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB096928; BAC76875.1; -.
SQ SEQUENCE 957 AA; 105646 MW; D92FB05E0CF7FEC0 CRC64;

Query Match 14.6%; Score 121.5; DB 11; Length 957;
Best Local Similarity 33.0%; Pred. No. 4.8e-05;
Matches 29; Conservative 10; Mismatches 38; Indels 11; Gaps 4;
```

```
QY 31 AAHLEGGKDNQFIWKPGRGCTGDCGGVQSVRAVWCFHVDGWTSHLSN---CGESNRP 87
DB 791 AIHPGCDKNCFAHWAQDWERCNTTCG-RGVKKLVLCMELANGKPOIRSGPEGLAKP 849
QY 88 PKERSCF-RVCDWHSDDLFWVEVSDWHHC 114
DB 850 PEESTCFERPC-----FKWYTPWSEC 871

RESULT 8
O95428 PRELIMINARY; PRT; 1235 AA.
AC O95428;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DE Hypothetical protein.
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RA Rowen L., Madan A., Qin S., Abbasi N., Dors M., Ratcliffe A.,
RA Madan A., Dickhoff R., Shafer R., Tasky S., Hood L.;
RT "Complete sequence of the gene for presentin 1.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF109907; AAC97963.1; -.
DR HSSP; P12111; 2KNT.
DR GO; GO:0004867; F:serine protease inhibitor activity; IEA.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_C2.
DR InterPro; IPR002223; Kunitz_BPTI.
DR InterPro; IPR000884; TSPI.
DR Pfam; PF00047; IG; 3.
DR Pfam; PF00014; Kunitz_BPTI; 1.
DR Pfam; PF00090; tsp_1; 5.
DR PRINTS; PR00759; BASICPTASE.
DR ProDom; PD000222; Kunitz_BPTI; 1.
DR SMART; SM00408; IGc2; 3.
DR SMART; SM00131; KU; 1.
DR SMART; SM00209; TSPI; 5.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 1.
DR PROSITE; PS0279; BPTI_KUNITZ_2; 1.
DR PROSITE; PS50835; IG LIKE; 3.
DR PROSITE; PS50092; TSPI; 5.
KW Hypothetical protein; Immunoglobulin domain; Protease inhibitor;
KW Serine protease inhibitor.
SQ SEQUENCE 1235 AA; 133477 MW; A0B44CCE4F38E350 CRC64;

Query Match 14.5%; Score 120.5; DB 4; Length 1235;
Best Local Similarity 37.5%; Pred. No. 8.2e-05;
Matches 27; Conservative 5; Mismatches 37; Indels 3; Gaps 3;

QY 44 WKPGRGCTGDCGGVQSVRAVWCFHVDGWTSHLSNCGESNRPKERSCFRV-CDWHS 102
DB 449 WSPFEPWGECSVGV-GVRKRVTCRGERGSLHTAACSLDRPPLTFCVHEDCPILLSD 507
QY 103 LFWVEVSDWHHC 114
DB 508 Q-AWHVGTWGLC 518

RESULT 9
O60345 PRELIMINARY; PRT; 951 AA.
AC O60345;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DE Hypothetical protein.
DE Hypothetical protein KIAA0605.
GN KIAA0605.
OS Homo sapiens (Human).
```

```
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RP TISSUE=Brain;
RX MEDLINE=98290545; PubMed=9628581;
RA Nagase T., Ishikawa K., Miyajima N., Tanaka A., Kotani H., Nomura N.,
RA Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:31-39(1998).
DR EMBL; AB011177; BAA25531.1; -.
DR PIR; T0260; T00260.
DR InterPro; IPR000884; TSPI.
DR Pfam; PF00090; tsp_1; 5.
DR SMART; SM00209; TSPI; 7.
DR PROSITE; PS50092; TSPI; 4.
KW Hypothetical protein.
SQ SEQUENCE 951 AA; 104678 MW; CC73753F232BC1A2 CRC64;

Query Match 14.4%; Score 119.5; DB 4; Length 951;
Best Local Similarity 34.1%; Pred. No. 8e-05;
Matches 30; Conservative 8; Mismatches 39; Indels 11; Gaps 4;

QY 31 AAHLEGGKDNQFIWKPGRGCTGDCGGVQSVRAVWCFHVDG---WTSLSNCGESNRP 87
DB 785 AIHPGCDKNCFAHWAQDWERCNTTCG-RGVKKLVLCMELANGKPOIRSGPEGLAKP 843
QY 88 PKERSCF-RVCDWHSDDLFWVEVSDWHHC 114
DB 844 PEESTCFERPC-----FKWYTPWSEC 865

RESULT 10
O86TH1 PRELIMINARY; PRT; 951 AA.
AC O86TH1;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DE Hypothetical protein KIAA0605.
DE Hypothetical protein KIAA0605.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RP TISSUE=Peripheral Nervous System;
RA Strausberg R.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC050544; AAH50544.1; -.
DR InterPro; IPR000884; TSPI.
DR Pfam; PF00090; tsp_1; 5.
DR SMART; SM00209; TSPI; 7.
DR PROSITE; PS50092; TSPI; 4.
KW Hypothetical protein.
SQ SEQUENCE 951 AA; 104620 MW; 93A9B0DC58AB6CC7 CRC64;

Query Match 14.4%; Score 119.5; DB 4; Length 951;
Best Local Similarity 34.1%; Pred. No. 8e-05;
Matches 30; Conservative 8; Mismatches 39; Indels 11; Gaps 4;

QY 31 AAHLEGGKDNQFIWKPGRGCTGDCGGVQSVRAVWCFHVDG---WTSLSNCGESNRP 87
DB 785 AIHPGCDKNCFAHWAQDWERCNTTCG-RGVKKLVLCMELANGKPOIRSGPEGLAKP 843
QY 88 PKERSCF-RVCDWHSDDLFWVEVSDWHHC 114
DB 844 PEESTCFERPC-----FKWYTPWSEC 865
```

```
RESULT 11
O15916 PRELIMINARY; PRT; 724 AA.
AC O15916;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Random slug cDNA22 protein.
GN RSC22.
OS Dictyostelium discoideum (slime mold).
OC Eukaryota; Mycetozoa; Dictyostelidae; Dictyostelium.
OC NCBI_TaxID=44683;
CX NCBI_TaxID=44683;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Iranfar N., Loomis W.F.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U82516; AAB69636.1; -.
DR InterPro; IPR009007; Rept_A_acid.
DR InterPro; IPR000884; TSP1.
DR SMART; SM00209; TSP1; 2.
DR PROSITE; PS50092; TSP1; 2.
SQ SEQUENCE 724 AA; 81684 MW; 0D9B2528F636A8B6 CRC64;

Query Match 14.2%; Score 118; DB 5; Length 724;
Best Local Similarity 34.7%; Pred. No. 8.9e-05;
Matches 25; Conservative 9; Mismatches 34; Indels 4; Gaps 4;

QY 44 WK-PGPGWRCCTGCGPGVQSRVAVWCFHVDGWTSHLSNCGESNRPKPKRSFCRVCDWHS 102
DB 512 WKVGNWNSDCQICG-SGLQTRVLCISINGELLDQERCNSHTKPPSERVC-AVAKCQAE 569

QY 103 LFQVEVSDWHHC 114
DB 570 -YHWDTSWMSG 580

RESULT 12
Q9UFZ4 PRELIMINARY; PRT; 898 AA.
AC Q9UFZ4;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Hypothetical protein (Fragment).
GN DKFZP434H204.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Wambutt R., Heubner D., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL110226; CAB53684.1; -.
DR PIR; T14764; T14764.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00090; tsp.1; 6.
DR SMART; SM00209; TSP1; 7.
DR PROSITE; PS50092; TSP1; 6.
DR KW Hypothetical protein.
FT NON TER
SQ SEQUENCE 898 AA; 96731 MW; 38532F1E56F0A11B CRC64;

Query Match 14.2%; Score 117.5; DB 4; Length 898;
Best Local Similarity 29.8%; Pred. No. 0.00013;
Matches 31; Conservative 4; Mismatches 42; Indels 27; Gaps 4;

QY 44 WKPGWRCCTGCGPGVQSRVAVWCFHVDGWTSHLS----NCGESNRPKPKRSFCR--VC 97
DB 95 WWAGEWLCSSCGPGGLSRRVAVLCIRSVGLDRQSALEPPFACEHLPRPTTETPCNRHVP 154
```

```
QY 98 DWHSDLFQVEVSDWHHCVLVYPARGVKPRTAECVTAQHGLQHR 141
DB 155 P-----ATWAVGNWSQC-----SVTCGGGTQRR 177

RESULT 13
Q91Z56 PRELIMINARY; PRT; 340 AA.
AC Q91Z56;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Similar to a disintegrin and metalloproteinase with thrombospondin
DE motifs 1 (ADAMTS-1) (Fragment).
GN ADAMTS15.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC009667; AAH09667.1; -.
DR MGD; MGI:2449569; Adams15.
DR CO; GO:0007223; P:integrin-mediated signaling pathway; IEA.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00090; tsp.1; 2.
DR SMART; SM00209; TSP1; 2.
DR PROSITE; PS50092; TSP1; 2.
DR KW Integrin.
FT NON TER
SQ SEQUENCE 340 AA; 37021 MW; 06562C747634A8BD CRC64;

Query Match 14.0%; Score 116.5; DB 11; Length 340;
Best Local Similarity 32.8%; Pred. No. 5.9e-05;
Matches 40; Conservative 16; Mismatches 51; Indels 15; Gaps 8;

QY 18 RKLFLLSLLSHAHLEGGKDNQ--FIWKPGRCTGCGPGVQSRVAVWCFHVDGWT 75
DB 206 RGLVLRLNSVLNSNQVE-QPDNPPARWAGVSGPCVSCG-SGLQRAVDCRDSFG-Q 262

QY 76 SHLSNCGESNRPKPKRSFCRVCDWHSDLFQVEVSDWHHCVLVYPARGVKPRTAECVTAQ 135
DB 263 QGASACDVDFHEPLEKACGEP--TWELGNWSPC-SKSCRG-FKERPLKCV--G 312

QY 136 HG 137
DB 313 HG 314

RESULT 14
Q9HBS6 PRELIMINARY; PRT; 237 AA.
AC Q9HBS6;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Gu J.R., Wan D.F., Zhao X.T., Zhou X.M., Jiang H.Q., Zhang P.P.,
RA Qin W.X., Huang Y., Qiu X.K., Qian L.F., He L.P., Li H.N., Yu Y.,
RA Yu J., Han L.H.;
RT "Novel Human cDNA clones with function of inhibiting cancer cell
RT growth."
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF217974; AAG17217.1; -.
DR Genew; HGNC:19706; TSR1.
```

	Query Match	14.0%;	Score 116;	DB 4;	Length 424;
	Best Local Similarity	26.6%;	Pred. No. 9.5e-05;		
	Matches	29;	Conservative	16;	Mismatches 54; Indels 10; Gaps 4
Qy	36	GKKNOFIIWKPGWGRCTGDCPGGVOSRAWVCFHVDGTSHL---	SNCGESNRNPKERS	92	
		:::::	:::		
Dd	256	GPCERTWRYYTGPWGECSECG-SGTQRDDIICVSKLGTFNTSPNSCHLPREP----	310		
		:::::	:::		
Qy	93	CPRVCWDHSDLPQWVSVDWHHCVLVPYARGVEKPTAECVTAQHQLHR	141		
		:::::	:::		
Dd	311	ALQPCCOGACADRWSTFWPSC--SRSCGGQTQTRVQCILSTR	357		
		:::::	:::		

Search completed: February 24, 2004, 01:07:27  
Job time : 12.692 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

M protein - protein search, using sw model

run on: February 24, 2004, 01:00:01 ; Search time 5.21344 Seconds  
(without alignments)  
1406.151 Million cell updates/sec

title: US-10-022-710-6\_COPY\_1\_142

effect score: 830

sequence: 1 MFPKSNLTVCWVRMSRKL.....EVKPRTAECVTAQHGLQHRM 142

coring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 389414 seqs, 51625971 residues

total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*

- 1: /cgn2\_6/prodata/2/iaa/5A-COMB.pep.\*
- 2: /cgn2\_6/prodata/2/iaa/5B-COMB.pep.\*
- 3: /cgn2\_6/prodata/2/iaa/6A-COMB.pep.\*
- 4: /cgn2\_6/prodata/2/iaa/6B-COMB.pep.\*
- 5: /cgn2\_6/prodata/2/iaa/PCITUS-COMB.pep.\*
- 6: /cgn2\_6/prodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	122	14.7	1882	4	US-09-369-364A-13
2	116.5	14.0	1745	4	US-09-800-729-89
3	114.5	13.8	1059	4	US-09-800-729-217
4	114	13.7	551	4	US-09-130-491-16
5	114	13.7	727	4	US-09-445-023A-12
6	109.5	13.2	1211	4	US-09-491-522-5
7	109.5	13.2	2165	4	US-09-800-729-155
8	106.5	12.8	1205	4	US-09-491-522-11
9	103	12.4	525	4	US-09-369-364A-21
10	100.5	12.1	1224	4	US-09-930-872-4
11	97.5	11.7	262	4	US-09-800-729-125
12	94	11.3	514	4	US-09-800-729-124
13	92	11.1	727	4	US-09-445-023A-1
14	92	11.1	949	4	US-09-568-559-2
15	92	11.1	967	4	US-09-130-491-2
16	88.5	10.7	450	4	US-09-369-364A-19
17	88.5	10.7	1081	4	US-09-369-364A-17
18	88	10.6	1184	2	US-08-918-914-1
19	88	10.6	1184	3	US-08-926-083-3
20	87	10.5	56	4	US-09-800-729-173
21	87	10.5	58	4	US-09-800-729-172
22	85.5	10.3	930	4	US-09-369-364A-2
23	85	10.2	142	4	US-09-380-882-8
24	84.5	10.2	59	4	US-09-800-729-163
25	83	10.0	481	4	US-09-130-491-8
26	83	10.0	481	4	US-09-452-991A-16955
27	81.5	9.8	518	4	US-09-369-364A-22

28	81.5	9.8	930	4	US-09-122-126B-15	Sequence 15, Appl
29	81.5	9.8	930	4	US-09-634-286A-15	Sequence 15, Appl
30	80	9.6	859	4	US-09-369-364A-5	Sequence 5, Appl
31	79.5	9.6	788	2	US-08-918-914-4	Sequence 4, Appl
32	78.5	9.5	608	4	US-09-130-491-13	Sequence 13, Appl
33	78	9.4	59	4	US-09-800-729-171	Sequence 171, Appl
34	77	9.3	757	4	US-09-963-791-24	Sequence 24, Appl
35	77	9.3	905	4	US-09-369-364A-9	Sequence 9, Appl
36	77	9.3	908	4	US-09-963-791-2	Sequence 2, Appl
37	76.5	9.2	143	3	US-09-012-084-2	Sequence 2, Appl
38	76.5	9.2	143	4	US-09-380-882-6	Sequence 6, Appl
39	76.5	9.2	144	1	US-07-974-848B-2	Sequence 2, Appl
40	76.5	9.2	837	4	US-09-122-126B-2	Sequence 2, Appl
41	76.5	9.2	837	4	US-09-634-286A-2	Sequence 2, Appl
42	74	8.9	144	4	US-09-180-100-21	Sequence 21, Appl
43	74	8.9	153	4	US-09-252-991A-26199	Sequence 26199, A
44	74	8.9	159	4	US-09-180-100-23	Sequence 23, Appl
45	74	8.9	376	4	US-09-180-100-22	Sequence 22, Appl

## ALIGNMENTS

## RESULT 1

US-09-369-364A-13  
; Sequence 13, Application US/09369364A  
; Patent No. 6391610  
; GENERAL INFORMATION:  
; APPLICANT: Apte, Suneel  
; APPLICANT: Hurskainen, Tiina L.  
; APPLICANT: Hirohata, Satoehi  
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases  
; FILE REFERENCE: 26473/4007/10-30-00  
; CURRENT APPLICATION NUMBER: US/09/369,364A  
; CURRENT FILING DATE: 1999-08-06  
; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 1882  
; TYPE: PRT  
; ORGANISM: Homo sapiens ADAMTS-9  
; FEATURE:  
; NAME/KEY: MOD\_RES  
; LOCATION: (468)  
; OTHER INFORMATION: Xaa = C  
; NAME/KEY: MOD\_RES  
; LOCATION: (521)  
; OTHER INFORMATION: Xaa = Y  
US-09-369-364A-13

Query Match 14.7%; Score 122; DB 4; Length 1882;  
Best Local Similarity 34.7%; Pred. No. 0.00015;  
Matches 25; Conservative 12; Mismatches 25; Indels 10; Gaps 4;  
QY 44 WKQPGWCTGDCPGQVQSVAVWCFHVDGWTSHLSCGSGNRPKRS-C-FRVCDWHS 102  
DB 1279 WRTGPGWACSTCA-GGQRRVVVQDENGTA--NDCVERIKPEQACSGFCP---- 1331

## RESULT 2

US-09-800-729-89  
; Sequence 89, Application US/09800729  
; Patent No. 6605392  
; GENERAL INFORMATION:  
; APPLICANT: Ni et al.  
; TITLE OF INVENTION: 32 Human secreted proteins  
; FILE REFERENCE: P204421  
; CURRENT APPLICATION NUMBER: US/09/800,729  
; CURRENT FILING DATE: 2001-03-08

;; PRIOR APPLICATION NUMBER: PCT/US00/26013  
;; PRIOR FILING DATE: 2000-09-22  
;; PRIOR APPLICATION NUMBER: 60/155,709  
;; PRIOR FILING DATE: 1999-09-24  
;; NUMBER OF SEQ ID NOS: 217  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 89  
;; LENGTH: 1745  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-09-800-729-89

Query Match 14.0%; Score 116.5; DB 4; Length 1745;  
Best Local Similarity 34.7%; Pred. No. 0.00054;  
Matches 33; Conservative 7; Mismatches 46; Indels 9; Gaps 5;  
QY 44 WKPGWGRCTGDC-GPG-GVQSRVAVCFHVDGWTSHLSNCGESNRPPKERSCF-RVCDWH 100  
DB 1593 WAFSSWGQCGPCIGPLAVQHRQVFCQTRDITLPSQCSALPRPVSTQNCWSEACSVH 1652  
QY 101 SDLQWEVSDWHHCVLVYPYARGEVKPRTACVTAQ 135  
DB 1653 -----WRVSLWTLCTATCGNYG-FQSRVVECVHAR 1681

RESULT 3  
US-09-800-729-217  
;; Sequence 217, Application US/09800729  
;; Patent No. 6605592  
;; GENERAL INFORMATION:  
;; APPLICANT: Ni et al.  
;; TITLE OF INVENTION: 32 Human secreted proteins  
;; FILE REFERENCE: P2044P1  
;; CURRENT APPLICATION NUMBER: US/09/800,729  
;; CURRENT FILING DATE: 2001-03-08  
;; PRIOR APPLICATION NUMBER: PCT/US00/26013  
;; PRIOR FILING DATE: 2000-09-22  
;; PRIOR APPLICATION NUMBER: 60/155,709  
;; PRIOR FILING DATE: 1999-09-24  
;; NUMBER OF SEQ ID NOS: 217  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 217  
;; LENGTH: 1059  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-09-800-729-217

Query Match 13.8%; Score 114.5; DB 4; Length 1059;  
Best Local Similarity 28.6%; Pred. No. 0.0005;  
Matches 34; Conservative 10; Mismatches 32; Indels 43; Gaps 8;  
QY 44 WKPGWGRCTGDCGPGVQSRVAVCFHVDGWTSHLSNCGESNRPPKERSCF----- 94  
DB 296 WEAGKWSKCTASCGQ-GVRRHVAC--VGG-----SDCDEGRPRQETTCYAGIPCSIAT 347  
QY 95 RVCDW-----HSDLFQW-----VSDWHHCVLVYPYARGEVKPRTACV 132  
DB 348 NSLDWMDRAYLDONTFGSDMNDND---WOAPRLVAGEWSTCSST--CGTGWMSRTVECV 401

RESULT 4  
US-09-130-491-16  
;; Sequence 16, Application US/09130491  
;; Patent No. 6416974  
;; GENERAL INFORMATION:  
;; APPLICANT: Holtzman, Douglas A.  
;; APPLICANT: Goodearl, Andrew D.J.  
;; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83  
;; FILE REFERENCE: 09404/041001  
;; CURRENT APPLICATION NUMBER: US/09/130,491  
;; CURRENT FILING DATE: 1998-08-07  
;; EARLIER APPLICATION NUMBER: US 60/058,108  
;; EARLIER FILING DATE: 1997-09-05

;; EARLIER APPLICATION NUMBER: US 60/054,961  
;; EARLIER FILING DATE: 1997-08-06  
;; NUMBER OF SEQ ID NOS: 16  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 16  
;; LENGTH: 551  
;; TYPE: PRT  
;; ORGANISM: Rattus rattus  
US-09-130-491-16

Query Match 13.7%; Score 114; DB 4; Length 551;  
Best Local Similarity 28.7%; Pred. No. 0.00027;  
Matches 37; Conservative 15; Mismatches 49; Indels 28; Gaps 8;  
QY 26 LLLSHA-----AHLEGKKNQF-----IWKPGWGRCTGDCGPGVQSRVAVCFHV 71  
DB 410 LMVGHALRPKIKFYFMKKTSTFNAIPTFSEWIEEWGECSTCG-SGWQRRVVQCRDI 468  
QY 72 DGWTSLSNCGESNRPPKERSCFRVCDWHSDL--FQWEVSDWHHCVLVYPYARGEVKPRTA 129  
DB 469 NGHPA--SECAKEVKFASTRPC-----ADLPCHPMQVGDWSPCKT-CGKG-YKKRTL 517  
QY 130 ECVTAQHGL 138  
DB 518 KCVSHDGGV 526

RESULT 5  
US-09-445-023A-12  
;; Sequence 12, Application US/09445023A  
;; Patent No. 6565859  
;; GENERAL INFORMATION:  
;; APPLICANT: Hirose, Kunitaka  
;; APPLICANT: Inoguchi, Ei-ji  
;; APPLICANT: Hakoizaki, Michinozi  
;; APPLICANT: Ishioaka, Keiko  
;; APPLICANT: Ishida, Yukako  
;; APPLICANT: Matsushima, Kouji  
;; APPLICANT: Kuno, Kouji  
;; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same, pharmaceutical  
;; TITLE OF INVENTION: composition and method of immunologically analyzing human ADAMTS  
;; FILE REFERENCE: Q57092  
;; CURRENT APPLICATION NUMBER: US/09/445,023A  
;; CURRENT FILING DATE: 1999-12-03  
;; PRIOR APPLICATION NUMBER: JP 9-160422  
;; PRIOR FILING DATE: 1997-06-03  
;; NUMBER OF SEQ ID NOS: 14  
;; SOFTWARE: PatentIn version 3.0  
;; SEQ ID NO 12  
;; LENGTH: 727  
;; TYPE: PRT  
;; ORGANISM: Mus sp.  
US-09-445-023A-12

Query Match 13.7%; Score 114; DB 4; Length 727;  
Best Local Similarity 28.7%; Pred. No. 0.00037;  
Matches 37; Conservative 15; Mismatches 49; Indels 28; Gaps 8;  
QY 26 LLLSHA-----AHLEGKKNQF-----IWKPGWGRCTGDCGPGVQSRVAVCFHV 71  
DB 586 LMVGHALRPKIKFYFMKKTSTFNAIPTFSEWIEEWGECSTCG-SGWQRRVVQCRDI 644  
QY 72 DGWTSLSNCGESNRPPKERSCFRVCDWHSDL--FQWEVSDWHHCVLVYPYARGEVKPRTA 129  
DB 645 NGHPA--SECAKEVKFASTRPC-----ADLPCHPMQVGDWSPCKT-CGKG-YKKRTL 693  
QY 130 ECVTAQHGL 138  
DB 694 KCVSHDGGV 702

RESULT 6  
US-09-491-522-5



Sequence 5, Application US/09491522

Patent No. 6428998

GENERAL INFORMATION:

APPLICANT: Collige, Alain

APPLICANT: Lapiere, Charles M.

APPLICANT: Prockop, Darwin J.

TITLE OF INVENTION: RECOMBINANT N-PROTEINASE,

TITLE OF INVENTION: AND THE PRODUCTION, METHODS AND USES THEREOF

NUMBER OF SEQUENCES: 17

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds, LLP

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: NY

COUNTRY: USA

ZIP: 10036-2811

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows

SOFTWARE: FastSeq for Windows Version 2.0b

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/491,522

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/886,333

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Abrams, Samuel B.

REGISTRATION NUMBER: 30,605

REFERENCE/DOCKET NUMBER: 8389-0060-999

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-493-4935

TELEFAX: 650-493-5556

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 1211 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-09-491-522-5

Query Match

Best Local Similarity 13.2%; Score 109.5; DB 4; Length 1211;

Matches 30; Conservative 12; Mismatches 38; Indels 27; Gaps 6;

40 NQFIWKPGWGRCTGCGGQVQSRVWCFHV--DGWTS--HLSNCGESNRPPKERSCFR 95

914 SQPVVWTGWEPCSCGTGRTGMQVRSVRCIQPLEHNTTSVHAKHCNDA-RPESRRACGR 972

96 -VCDWHSDLQWEVSDWHHCVLVPYARGEVKPRTASCVTAGHGLQHR 141

973 ELCPG-----RWRAGPWSQC-----SVTCGNGTQER 998

RESULT 7

US-09-800-729-155

Sequence 155, Application US/09800729

Patent No. 6605592

GENERAL INFORMATION:

APPLICANT: Ni et al.

TITLE OF INVENTION: 32 Human secreted proteins

FILE REFERENCE: P2044P1

CURRENT APPLICATION NUMBER: US/09/800,729

CURRENT FILING DATE: 2001-03-08

PRIOR APPLICATION NUMBER: PCT/US00/26013

PRIOR FILING DATE: 2000-09-22

PRIOR APPLICATION NUMBER: 60/155,709

PRIOR FILING DATE: 1999-09-24

NUMBER OF SEQ ID NOS: 217

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 155

LENGTH: 2165

TYPE: PRT

ORGANISM: Homo sapiens

US-09-800-729-155

Query Match

Best Local Similarity 13.2%; Score 109.5; DB 4; Length 2165;

Matches 28; Conservative 13; Mismatches 32; Indels 25; Gaps 4;

1 MFPSNLTVTCTWVRSVRKLFLLLSLHAAHLEKGNQFIWKPGWGRCTGDCGPGG 60

1865 LFPAPNESQTCEL-----NPCDSEFKWSTGFWGECNKCQ-G 1901

61 VQSRVWCFHVDGWTSHLSNCGESNRPPKERSCF-RVC 97

1902 IRRRVKCVANDGRRVERVKC-TTKKPRTOYCFERNC 1938

RESULT 8

US-09-491-522-11

Sequence 11, Application US/09491522

Patent No. 6428998

GENERAL INFORMATION:

APPLICANT: Collige, Alain

APPLICANT: Lapiere, Charles M.

APPLICANT: Prockop, Darwin J.

TITLE OF INVENTION: RECOMBINANT N-PROTEINASE,

TITLE OF INVENTION: AND THE PRODUCTION, METHODS AND USES THEREOF

NUMBER OF SEQUENCES: 17

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds, LLP

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: NY

COUNTRY: USA

ZIP: 10036-2811

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows

SOFTWARE: FastSeq for Windows Version 2.0b

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/491,522

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/886,333

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Abrams, Samuel B.

REGISTRATION NUMBER: 30,605

REFERENCE/DOCKET NUMBER: 8389-0060-999

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-493-4935

TELEFAX: 650-493-5556

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:

LENGTH: 1205 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-09-491-522-11

Query Match

Best Local Similarity 12.8%; Score 106.5; DB 4; Length 1205;

Matches 30; Conservative 13; Mismatches 37; Indels 27; Gaps 6;

40 NQFIWKPGWGRCTGDCGPGVQSRVWCFHV--FHVDGWTSHLSNCGESNRPPKERSCFR 95

908 SQPVVWTGWEVFCRSRGRTGMQVRSVRCVQLHNTTTSVHTKHCNDA-RPEGRACNR 966

```
QY 96 -VCDWHSDFQWVSDHHCVCVLPYARGVKPRTAEVCVTAQHGLOHR 141
Db 967 ELCPG-----RWAGSQSC-----SVTCNGTQER 992

RESULT 9
US-09-369-364A-21
; Sequence 21, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Homo sapiens ADAMTS-R1
US-09-369-364A-21

Query Match 12.4%; Score 103; DB 4; Length 525;
Best Local Similarity 25.2%; Pred. No. 0.004;
Matches 28; Conservative 14; Mismatches 27; Indels 42; Gaps 6;

QY 44 WKPGPWGRTGDCGGVQSVRAVWCFF-----VDGWTSHLNCGESNRPPKERSCF 94
Db 380 WEATPTWTCSSCG-GGQSRVSCVEEDIQHVTSVSEW-----KMYTPKMPVIAQPC- 432

QY 95 RYCDWHSDF-----QWVSDWHHCVCVLPYARGVKPRTAEVCVTAQHGLOHR 142
Db 433 -----NIFDCPKLAQEWSPCT-----VTCGGGLRYRV 460

RESULT 10
US-09-930-872-4
; Sequence 4, Application US/09930872
; Patent No. 6448388
; GENERAL INFORMATION:
; APPLICANT: Fridde, Carl Johan
; APPLICANT: Hilbun, Erin
; TITLE OF INVENTION: No. 6448388el Human Proteases and Polynucleotides Encoding the Sa
; FILE REFERENCE: LEX-0219-USA
; CURRENT APPLICATION NUMBER: US/09/930,872
; CURRENT FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: US 60/225,852
; PRIOR FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1224
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-930-872-4

Query Match 12.1%; Score 100.5; DB 4; Length 1224;
Best Local Similarity 28.0%; Pred. No. 0.002;
Matches 30; Conservative 6; Mismatches 28; Indels 43; Gaps 6;

QY 44 WKPGPWGRTGDCGGVQSVRAVWCFFHVDGWTSHLNCGESNRPPKERSCFVCDWH--- 100
Db 1130 WFAFSPHSQCTASCG-GGVQTRSVQCL---AGGRPASGCLLHKQP-----SASLACNTHFCP 1181

QY 101 -----SDLPQWVSDWHHCVCVLPYARGVKPRTAEVCVTAQHGL 138
Db 1182 IAEKDAFCNDYFHW-----CYLVP-----QHGML 1205

RESULT 11
```

```
US-09-800-729-125
; Sequence 125, Application US/09800729
; Patent No. 6605592
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 32 Human secreted proteins
; FILE REFERENCE: P2044P1
; CURRENT APPLICATION NUMBER: US/09/800,729
; CURRENT FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: PCT/US00/26013
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: 60/155,709
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 217
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 125
; LENGTH: 262
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (254)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-800-729-125

Query Match 11.7%; Score 97.5; DB 4; Length 262;
Best Local Similarity 29.4%; Pred. No. 0.0073;
Matches 40; Conservative 16; Mismatches 55; Indels 25; Gaps 10;

QY 20 LFLALLSLLSHAHLEGGKDNQFIWKP-GPWGRCCTGDCGGVQSV--RAVWCFFHYDGMTS 76
Db 14 LFLAFLLSSRTARSEEDRDG--LWDANGFWSECSRTCGGASYSLLRCLSSKSCGRNI 71

QY 77 HLSNCGESNRPPKERSCFRV--CDWHSDF-----FQW-EVSD--WHHCVCVLPYARG--- 122
Db 72 RYRTCSNVDCPP-EAGDFRAQCSAHNDYKHGQFYEMLPVSDNDPNPCSLKQAKGTTL 130

QY 123 --EVKPRTAEE---CVT 133
Db 131 VVELAPKVLDTGTRCYT 146

RESULT 12
US-09-800-729-124
; Sequence 124, Application US/09800729
; Patent No. 6605592
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 32 Human secreted proteins
; FILE REFERENCE: P2044P1
; CURRENT APPLICATION NUMBER: US/09/800,729
; CURRENT FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: PCT/US00/26013
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: 60/155,709
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 217
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 124
; LENGTH: 514
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-800-729-124

Query Match 11.3%; Score 94; DB 4; Length 514;
Best Local Similarity 33.3%; Pred. No. 0.038;
Matches 25; Conservative 6; Mismatches 34; Indels 10; Gaps 4;

QY 44 WKPGPWGRTGDCGGVQSVRAVWCFF--VDGWTSHLSN--CGESNRPPKERSCFVCDW 99
Db 23 WYFAQWQPCRTCG-GGVQKREVLCKQRVADGSLFELPETFCSAS-----KPACQQAACK 76

QY 100 HSDLFQWVSDWHHC 114
```



GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

MM protein - protein search, using sw model

run on: February 24, 2004, 01:07:37 ; Search time 59.0268 Seconds  
(without alignments)  
507.969 Million cell updates/sec

Title: US-10-022-710-6\_COPY\_1\_142

Perfect score: 830

Sequence: 1 MFPEKSLNLTVCWVRMRKL.....EVKPTAEVTAQHGLQHRM 142

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 21153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	573	69.0	247	9	US-09-864-761-43221
2	327.5	39.5	933	15	Sequence 2928, Ap
3	289	34.8	277	9	Sequence 46564, A
4	252.5	30.4	1588	13	Sequence 2, Appli
5	252.5	30.4	1588	15	Sequence 25, Appli
6	252.5	30.4	1588	15	Sequence 41, Appli
7	128.5	15.5	1057	14	Sequence 10, Appli
8	128.5	15.5	1122	14	Sequence 13, Appli
9	128.5	15.5	1145	14	Sequence 20, Appli
10	127	15.3	1629	9	US-09-972-467-2
11	127	15.3	1907	9	Sequence 2, Appli
12	127	15.3	1916	15	US-10-938-330-25
13	126	15.2	1280	14	Sequence 10, Appli
14	122	14.7	1882	9	Sequence 86, Appli
15	121.5	14.6	226	9	US-09-918-171A-13
					Sequence 8, Appli

Sequence 12, Appli  
Sequence 6, Appli  
Sequence 10, Appli  
Sequence 4, Appli  
Sequence 8, Appli  
Sequence 2, Appli  
Sequence 124, Appli  
Sequence 6, Appli  
Sequence 47, Appli  
Sequence 2, Appli  
Sequence 5, Appli  
Sequence 762, Appli  
Sequence 89, Appli  
Sequence 222, Appli  
Sequence 2, Appli  
Sequence 554, Appli  
Sequence 45, Appli  
Sequence 10, Appli  
Sequence 10, Appli  
Sequence 10, Appli  
Sequence 217, Appli  
Sequence 10, Appli  
Sequence 16, Appli  
Sequence 16, Appli  
Sequence 12, Appli  
Sequence 12, Appli  
Sequence 12, Appli  
Sequence 4, Appli  
Sequence 7, Appli  
Sequence 5, Appli

ALIGNMENTS

RESULT 1

US-09-864-761-43221  
; Sequence 43221, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aecmca-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30





Matches 39; Conservative 21; Mismatches 24; Indels 3; Gaps 2;  
QY 55 DCGPGVQSRVAVCFHVDGWTSHLNCGESNRPPKERSCFRVCDSHSLDFQWEVSDWHHC 114  
Db 4 ECGPGIGIQTAVCAHVEGWTILHFNCKQAEPPNNQCNCFKVCDSHSLDFQWEVSDWHHC 63  
QY 115 VLVVPYARGVKPRTAECVTAQGLQHR 141  
Db 64 QPV-ISKSLKP--LECIKGEGIQVR 87

RESULT 7  
US-10-188-869-10  
; Sequence 10, Application US/10188869  
; Publication No. US20030148306A1  
; GENERAL INFORMATION:  
; APPLICANT: LAVALLIE, EDWARD  
; APPLICANT: RACIE, LISA  
; APPLICANT: DIBLASIO, ELIZABETH  
; APPLICANT: AGOSTINO, MICHAEL  
; TITLE OF INVENTION: AGGREGANASE MOLECULES  
; FILE REFERENCE: 08702.0092-00000  
; CURRENT APPLICATION NUMBER: US/10/188,869  
; CURRENT FILING DATE: 2002-07-05  
; PRIOR APPLICATION NUMBER: 60/349,133  
; PRIOR FILING DATE: 2002-01-16  
; PRIOR APPLICATION NUMBER: 60/303,051  
; PRIOR FILING DATE: 2001-06-05  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 10  
; LENGTH: 1057  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-188-869-10

Query Match 15.5%; Score 128.5; DB 14; Length 1057;  
Best Local Similarity 31.1%; Pred. No. 0.00014;  
Matches 33; Conservative 16; Mismatches 40; Indels 17; Gaps 7;  
QY 35 EGKDNQFTWKPGWRCRTGDCGPGVQSRVAVCFHVDGWTSHLNCGESNRPPKERSCF 94  
Db 879 EG-QDCLSIWEASEWSQCSASCGK-GWVKRTVAC-----TNSQKCDASTRPAEEA-- 928  
QY 95 RVCDSHSLDFQWEVSDWHHCVLVPYARGVKPRTAEC---VTAQHG 137  
Db 929 --CEDYSGCYEWKTDWSTCSST-CGKG-LQSRVVCQMHKVTGRHG 970

RESULT 8  
US-10-188-869-13  
; Sequence 13, Application US/10188869  
; Publication No. US20030148306A1  
; GENERAL INFORMATION:  
; APPLICANT: LAVALLIE, EDWARD  
; APPLICANT: RACIE, LISA  
; APPLICANT: DIBLASIO, ELIZABETH  
; APPLICANT: AGOSTINO, MICHAEL  
; TITLE OF INVENTION: AGGREGANASE MOLECULES  
; FILE REFERENCE: 08702.0092-00000  
; CURRENT APPLICATION NUMBER: US/10/188,869  
; CURRENT FILING DATE: 2002-07-05  
; PRIOR APPLICATION NUMBER: 60/349,133  
; PRIOR FILING DATE: 2002-01-16  
; PRIOR APPLICATION NUMBER: 60/303,051  
; PRIOR FILING DATE: 2001-06-05  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 1122  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-188-869-13

Query Match 15.5%; Score 128.5; DB 14; Length 1122;  
Best Local Similarity 31.1%; Pred. No. 0.00015;  
Matches 33; Conservative 16; Mismatches 40; Indels 17; Gaps 7;  
QY 35 EGKDNQFTWKPGWRCRTGDCGPGVQSRVAVCFHVDGWTSHLNCGESNRPPKERSCF 94  
Db 944 EG-QDCLSIWEASEWSQCSASCGK-GWVKRTVAC-----TNSQKCDASTRPAEEA-- 993  
QY 95 RVCDSHSLDFQWEVSDWHHCVLVPYARGVKPRTAEC---VTAQHG 137  
Db 994 --CEDYSGCYEWKTDWSTCSST-CGKG-LQSRVVCQMHKVTGRHG 1035

RESULT 9  
US-10-188-869-20  
; Sequence 20, Application US/10188869  
; Publication No. US20030148306A1  
; GENERAL INFORMATION:  
; APPLICANT: LAVALLIE, EDWARD  
; APPLICANT: RACIE, LISA  
; APPLICANT: DIBLASIO, ELIZABETH  
; APPLICANT: AGOSTINO, MICHAEL  
; TITLE OF INVENTION: AGGREGANASE MOLECULES  
; FILE REFERENCE: 08702.0092-00000  
; CURRENT APPLICATION NUMBER: US/10/188,869  
; CURRENT FILING DATE: 2002-07-05  
; PRIOR APPLICATION NUMBER: 60/349,133  
; PRIOR FILING DATE: 2002-01-16  
; PRIOR APPLICATION NUMBER: 60/303,051  
; PRIOR FILING DATE: 2001-06-05  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 20  
; LENGTH: 1145  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-188-869-20

Query Match 15.5%; Score 128.5; DB 14; Length 1145;  
Best Local Similarity 31.1%; Pred. No. 0.00016;  
Matches 33; Conservative 16; Mismatches 40; Indels 17; Gaps 7;  
QY 35 EGKDNQFTWKPGWRCRTGDCGPGVQSRVAVCFHVDGWTSHLNCGESNRPPKERSCF 94  
Db 967 EG-QDCLSIWEASEWSQCSASCGK-GWVKRTVAC-----TNSQKCDASTRPAEEA-- 1016  
QY 95 RVCDSHSLDFQWEVSDWHHCVLVPYARGVKPRTAEC---VTAQHG 137  
Db 1017 --CEDYSGCYEWKTDWSTCSST-CGKG-LQSRVVCQMHKVTGRHG 1058

RESULT 10  
US-09-972-467-2  
; Sequence 2, Application US/09972467  
; Patent No. US20020090373A1  
; GENERAL INFORMATION:  
; APPLICANT: PFIZER INC.  
; TITLE OF INVENTION: ADAMTS POLYPEPTIDES, NUCLEIC ACIDS ENCODING THEM, AND  
; FILE REFERENCE: PCL0850A  
; CURRENT APPLICATION NUMBER: US/09/972,467  
; CURRENT FILING DATE: 2001-10-05  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 1629  
; TYPE: PRT  
; ORGANISM: Human  
US-09-972-467-2

Query Match 15.3%; Score 127; DB 9; Length 1629;  
Best Local Similarity 31.5%; Pred. No. 0.00032;



Matches 28; Conservative 15; Mismatches 34; Indels 12; Gaps 5;  
2Y 44 WKPGMGRCTGCGPGVQSVAYWCFHVDGWTSHLSNCGESNRPPKERSG-FRVCDWHS 102  
1332 WRTGPGWACSTCA-GGSQRRVVVCQDENGTYA--NDCVERIKPDQRACSGPCP---- 1384  
2Y 103 LFWQVSDWHCHVLPVYARGVPRTAEC 131  
1385 --QWAGNMGECTKL--CGGIRTRLVVC 1409  
RESULT 11  
US-09-938-330-25  
; Sequence 25, Application US/09938330  
; Patent No. US20020115838A1  
; GENERAL INFORMATION:  
; APPLICANT: Walke, D. Wade  
; APPLICANT: Hilbun, Erin  
; APPLICANT: Scoville, John  
; APPLICANT: Friddle, Carl Johan  
; APPLICANT: Hu, Yi  
; APPLICANT: Turner, C. Alexander Jr.  
; TITLE OF INVENTION: No. US20020115838A1 Human Proteases and Polynucleotides Encodin  
; FILE REFERENCE: LEX-0237-USA  
; CURRENT APPLICATION NUMBER: US/09/938,330  
; PRIOR FILING DATE: 2001-08-22  
; PRIOR APPLICATION NUMBER: US 60/227,104  
; PRIOR FILING DATE: 2000-08-22  
; PRIOR APPLICATION NUMBER: US 60/233,796  
; PRIOR FILING DATE: 2000-09-19  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: Fast-Seq for Windows Version 4.0  
; SEQ ID NO 25  
; LENGTH: 1907  
; TYPE: PRT  
; ORGANISM: homo sapiens  
JS-09-938-330-25  
Query Match 15.3%; Score 127; DB 9; Length 1907;  
Best Local Similarity 31.5%; Pred. No. 0.00038;  
Matches 28; Conservative 15; Mismatches 34; Indels 12; Gaps 5;  
2Y 44 WKPGMGRCTGCGPGVQSVAYWCFHVDGWTSHLSNCGESNRPPKERSG-FRVCDWHS 102  
1304 WRTGPGWACSTCA-GGSQRRVVVCQDENGTYA--NDCVERIKPDQRACSGPCP---- 1356  
2Y 103 LFWQVSDWHCHVLPVYARGVPRTAEC 131  
1357 --QWAGNMGECTKL--CGGIRTRLVVC 1381  
RESULT 12  
US-10-274-639-10  
; Sequence 10, Application US/10274639  
; Publication No. US20030232349A1  
; GENERAL INFORMATION:  
; APPLICANT: INCYTE GENOMICS, INC.  
; APPLICANT: DELEGANE, Angelo M.; GANDHI, Ameena R.  
; APPLICANT: HAPALIA, April J.A.; LU, Dyung Aina M.  
; APPLICANT: PATTERSON, Chandra; TRIBOULEY, Catherine M.  
; APPLICANT: DAS, Debopriya; KALLICK, Deborah A.  
; APPLICANT: NGUYEN, Darnell B.; LEE, Ernestine A.  
; APPLICANT: KHAN, Farrah A.; YUE, Henry  
; APPLICANT: AU-YOUNG, Janice K.; GRIFFIN, Jennifer A.  
; APPLICANT: POLICKY, Jennifer L.; RAMKUMAR, Jayalaxmi  
; APPLICANT: YANG, Junming; THANGAVELU, Kavitha  
; APPLICANT: DING, Li; KEARNEY, Liam  
; APPLICANT: BAUGHN, Mariah R.; BOROWSKY, Mark L.  
; APPLICANT: SANJANWALA, Madhusudan M.; YAO, Monique G.  
; APPLICANT: BURFORD, Neil; WALLIA, Narinder K.  
; APPLICANT: LAL, Preeti G.; LEE, Sally  
; APPLICANT: TODD, Stephen; LO, Terence P.  
; APPLICANT: TANG, Y. Tom; ELLIOTT, Vicki S.

; APPLICANT: AZIMZAI, Yalda; LU, Yan  
; TITLE OF INVENTION: PROTEASES  
; FILE REFERENCE: PI-0167 USA  
; CURRENT APPLICATION NUMBER: US/10/274,639  
; CURRENT FILING DATE: 2002-10-18  
; PRIOR APPLICATION NUMBER: PCT/US01/22397  
; PRIOR FILING DATE: 2001-07-17  
; PRIOR APPLICATION NUMBER: US 60/220,063  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: US 60/221,680  
; PRIOR FILING DATE: 2000-07-28  
; PRIOR APPLICATION NUMBER: US 60/223,544  
; PRIOR FILING DATE: 2000-08-04  
; PRIOR APPLICATION NUMBER: US 60/224,717  
; PRIOR FILING DATE: 2000-08-11  
; PRIOR APPLICATION NUMBER: US 60/225,988  
; PRIOR FILING DATE: 2000-08-16  
; PRIOR APPLICATION NUMBER: US 60/227,568  
; PRIOR FILING DATE: 2000-08-23  
; NUMBER OF SEQ ID NOS: 42  
; SOFTWARE: PERL Program  
; SEQ ID NO 10  
; LENGTH: 1916  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Incyte ID No. US20030232349A1 2994162CD1  
US-10-274-639-10  
Query Match 15.3%; Score 127; DB 15; Length 1916;  
Best Local Similarity 31.5%; Pred. No. 0.00038;  
Matches 28; Conservative 15; Mismatches 34; Indels 12; Gaps 5;  
QY 44 WKPGMGRCTGCGPGVQSVAYWCFHVDGWTSHLSNCGESNRPPKERSG-FRVCDWHS 102  
1313 WRTGPGWACSTCA-GGSQRRVVVCQDENGTYA--NDCVERIKPDQRACSGPCP---- 1365  
QY 103 LFWQVSDWHCHVLPVYARGVPRTAEC 131  
1366 --QWAGNMGECTKL--CGGIRTRLVVC 1390  
RESULT 13  
US-10-087-887-86  
; Sequence 86, Application US/10087887  
; Publication No. US20030198957A1  
; GENERAL INFORMATION:  
; APPLICANT: Kekuda, Ramesh  
; APPLICANT: Conley, Pamela B.  
; APPLICANT: Yang, Ruey-Bing  
; APPLICANT: Hart, Matthew  
; APPLICANT: Tomlinson, James E.  
; APPLICANT: Topper, James N.  
; APPLICANT: Shimkets, Richard A.  
; APPLICANT: Leach, Martin D.  
; APPLICANT: Zethusen, Bryan D.  
; APPLICANT: Komuves, Laszlo  
; APPLICANT: Padigaru, Muralidhara  
; TITLE OF INVENTION: PROTEINS AND NUCLEIC ACIDS ENCODING SAME  
; FILE REFERENCE: 21402-285  
; CURRENT APPLICATION NUMBER: US/10/087,887  
; CURRENT FILING DATE: 2002-03-01  
; PRIOR APPLICATION NUMBER: 60/273,049  
; PRIOR FILING DATE: 2001-03-02  
; PRIOR APPLICATION NUMBER: 60/279,883  
; PRIOR FILING DATE: 2001-03-29  
; PRIOR APPLICATION NUMBER: 60/277,791  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/281,248  
; PRIOR FILING DATE: 2001-04-03  
; PRIOR APPLICATION NUMBER: 60/282,864  
; PRIOR FILING DATE: 2001-04-10

; PRIOR APPLICATION NUMBER: 60/282,537  
; PRIOR FILING DATE: 2001-04-09  
; PRIOR APPLICATION NUMBER: 60/282,867  
; PRIOR FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 104  
; SOFTWARE: CuraseqList version 0.1  
; SEQ ID NO 86  
; LENGTH: 1280  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-10-087-887-86

Query Match 15.2%; Score 126; DB 14; Length 1280;  
Best Local Similarity 32.6%; Pred. No. 0.00032;  
Matches 31; Conservative 15; Mismatches 39; Indels 10; Gaps 6;  
QY 42 FIWPGWGRCTGDCGGVQSVRAVMC-FHVDGWTSHLSNCGESNRPPKERSC-FRVCDW 99  
Db 307 FSWSHSGWGDCAECG-GGHQSLVFTIDNEAYPDHM--CQHQPRPTHRSCNTQPC-- 361  
QY 100 HSDLFQWEVSDWHHCVLVYPYARGVKPRTACVTA 134  
Db 362 -PKTKRWKVPWTPCSV--SCGGVQSVSVYCISS 393

## RESULT 14

US-09-918-171A-13  
; Sequence 13, Application US/09918171A  
; Patent No. US20020110894A1  
; GENERAL INFORMATION:  
; APPLICANT: Apte, Suneel  
; APPLICANT: Hurskainen, Tiina L.  
; APPLICANT: Hirohata, Satoshi  
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases  
; FILE REFERENCE: 26473/04193  
; CURRENT APPLICATION NUMBER: US/09/918,171A  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: 09/369,364  
; PRIOR FILING DATE: 1999-08-06  
; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 1882  
; TYPE: PRT  
; ORGANISM: Homo sapiens ADAMTS-9  
; FEATURE:  
; NAME/KEY: MOD RES  
; LOCATION: (468)  
; OTHER INFORMATION: xaa = Cys  
; NAME/KEY: MOD RES  
; LOCATION: (521)  
; OTHER INFORMATION: xaa = Tyr  
US-09-918-171A-13

Query Match 14.7%; Score 122; DB 9; Length 1882;  
Best Local Similarity 34.7%; Pred. No. 0.0012;  
Matches 25; Conservative 12; Mismatches 25; Indels 10; Gaps 4;  
QY 44 WKPGWGRCTGDCGGVQSVRAVMCFHVDGWTSHLSNCGESNRPPKERSC-FRVCDWHS 102  
Db 1279 WRTGPGACSSITCA-GGSQRRVVVQDENGTYA--NDCVERIKPDEQACSGPCP----- 1331  
QY 103 LFQWEVSDWHHC 114  
Db 1332 --QWAYGNWGE 1341

## RESULT 15

US-09-945-676-8  
; Sequence 8, Application US/09945676  
; Patent No. US20020077465A1  
; GENERAL INFORMATION:  
; APPLICANT: Shi et al.

; TITLE OF INVENTION: ADAM Polynucleotides, Polypeptides, and Antibodies  
; FILE REFERENCE: PT051P1  
; CURRENT APPLICATION NUMBER: US/09/945,676  
; CURRENT FILING DATE: 2001-09-05  
; PRIOR APPLICATION NUMBER: PCT/US01/05497  
; PRIOR FILING DATE: 2001-02-22  
; PRIOR APPLICATION NUMBER: 60/187,937  
; PRIOR FILING DATE: 2000-03-03  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 226  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-945-676-8

Query Match 14.6%; Score 121.5; DB 9; Length 226;  
Best Local Similarity 28.2%; Pred. No. 0.00015;  
Matches 3; Conservative 19; Mismatches 43; Indels 17; Gaps 7;  
QY 31 AAHLEGGKDNQFTWKPGWGRCTGDCGGVQSVRAVMCFHVDGWTSHLSNCGESNRPPKE 90  
Db 47 AQRCGE-QDCMTVWEAGVWSECSVKCGK-GIRHRTVRC-----TNPRKKCVLSTPREA 98  
QY 91 RSCFRVCDWHSDDLQWEVSDWHHCVLVYPYARGVKPRTAEC---VTAQHG 137  
Db 99 ED----CEDYSKCYVVRMGDWSKC-SITCGKG-MQSRVQCVMHKITGRHG 142

Search completed: February 24, 2004, 01:30:05  
Job time : 59.0268 secs